Lectures Notes : Algorithmic Information Theory

Itisan Halias

itha.mail0@gmail.com

September 2025

Contents

Chapter 1

Algorithmic "Solomonoff" Probability

1.1 Semimeasures

1.1.1 Definitions

Definition: Measure

A function $\mu : \mathbb{B}^* \to \mathbb{R}$ is a measure if

1. $\mu(\varepsilon) = 1$,
2. $\mu(x) = \sum_{b \in \mathbb{B}} \mu(xb)$ for all $x \in \mathbb{B}^*$.

Definition-Property: Uniform or Lebesgue measure

Let $x \in \mathbb{B}^*$, we define the uniform measure as

$$\lambda(x) = 2^{-\ell(x)}$$

Definition: Semimeasures

A function $\mu : \mathbb{B}^* \to \mathbb{R}$ is a *semimeasure* if

1. $\mu(\varepsilon) \leq 1$,
2. $\mu(x) \geq \mu(x0) + \mu(x1)$ for all $x \in \mathbb{B}^*$.

Definition: Lower semicomputable semimeasures

A semimeasure μ is called lower semicomputable if μ is lower semicomputable. We denote by Π_{sm} the set of lower semicomputable semimeasures.

1.1.2 Enumeration of lower semicomputable semimeasures

Algorithm: Partial mapping $\langle\, \phi^*_{lc} \,\rangle \xrightarrow{\Pi_{\nearrow} \Rightarrow \Pi^{\sim}_{sm}} \langle\, \nu_i \,\rangle$

Let ϕ^*_{lc} be a strict lower approximation of a semimeasure. Let us define, for each integer $i \geq 0$, the computable function ν_i operating as follows:

Input : $x \in \mathbb{B}^*$

Step 1. For each $p \in \mathbb{B}^{=i}$ do:

- If $1 \geq \phi_{lc}(p, i)$ then
  - Assign $K^i_p \leftarrow i$. [†]
- Else: Loop forever.

Step 2. For each $j = i-1, i-2, \ldots, 1, 0$ do: [loop 1]

- For each p in $\mathbb{B}^{=j}$ do: [loop 2]
  - By enumerating $\phi_{lc}(p,0), \phi_{lc}(p,1), \ldots$ find the first (that is, the smallest) $K \in \mathcal{N}$ satisfying the conditions opposite:

$$K^i_p \leftarrow \min\left\{ K \;\middle|\; \begin{array}{l} K \geq i \quad \text{and} \\ 1 \geq \phi_{lc}\big(p, K\big) \geq \phi_{lc}(p1, K^i_{p1}) + \phi_{lc}(p0, K^i_{p0}) \end{array} \right\} \quad [\sharp]$$

    (Note that if such a K does not exist then the enumeration does not terminate, thus the procedure loops forever)

Step 3. If K^i_x is not assigned:

- Write 0 as output and accept.

Else:

- Write $\phi_{lc}(x, K^i_x)$ as output and accept.

Remark: For a given approximation ϕ^*_{lc} and integer $i \geq 0$. The values K^i_p do not depend on the input: The calls $\nu_i(x)$ and $\nu_i(x')$ assign the same $(K^i_p)_{p \in \mathbb{B}^*}$. We define $K^i_p = \varnothing$ if the variable K^i_p is not assigned in $\nu_i(\cdot)$.

In the three following lemmas, we define $(\nu_i)_{i \in \mathcal{N}}$ as a sequence of partially computable functions as defined by the algorithm below for a given strict lower approximation ϕ^*_{lc}.

Lemma: Analysis of M_{map}

For all $i \in \mathcal{N}$ and $p \in \mathbb{B}^*$ we have:

1. $|p| > i \implies K^i_p = \varnothing$.
2. $K^i_p \neq \varnothing \implies \forall p'$ satisfying $i \geq |p'| > |p|$ we have $K^i_{p'} \neq \varnothing$.
3. $K^{i+1}_p \neq \varnothing \implies K^i_p \neq \varnothing$ and $K^{i+1}_p \geq K^i_p$.
4. $K^i_x \neq \varnothing$ and $x \in dom(\nu_i) \implies 1 \geq \nu_i(x)$.

Proof. Let $p \in \mathbb{B}^*$ and $i \in \mathcal{N}$

1) The K_p^i are assigned either at [†] or at [♯]. In both cases $|p| \leq i$, thus we necessarily have that if $|p| > i$ then $K_p^i = \varnothing$.

2) Assume that $K_p^i \neq \varnothing$.

- Case $|p| = i$: This case is accounted for by hypothesis.
- Case $|p| < i$: For $K_p^i \neq \varnothing$, it is necessary that, in loop 1, we had $j = |p|$. Now, as this loop is decreasing, we necessarily obtain that the assignment in [♯] of the $K_{p'}^i$, for $i > |p'| > |p|$, succeeded, that is $K_{p'} \neq \varnothing$.

3) Let $i \in \mathcal{N}$. Let us proceed by downward induction on P_n with $n \in \{0, \ldots, i\}$:

$$P_n : \text{"For all } p \in \mathbb{B}^{=n}\text{, if } K_p^{i+1} \neq \varnothing \text{ then } K_p^i \neq \varnothing \text{ and } K_p^{i+1} \geq K_p^i\text{"}$$

- Base case ($n = i$): Assume $K_p^{i+1} \neq \varnothing$ for a $p \in \mathbb{B}^{=i}$. The variable K_p^{i+1} is assigned by hypothesis, which by ♯ implies that $K_p^{i+1} \geq i+1$ and $1 \geq \phi_{lc}^*(p, K_p^{i+1})$.

  Moreover, due to the fact that ϕ_{lc}^* is a lower approximation, we have

  $$\phi_{lc}^*(p, i) \leq \phi_{lc}^*(p, K_p^{i+1}) \leq 1$$

  Thus by step 1, we note that $K_p^i = i$ will be assigned because the condition $1 \geq \phi_{lc}^*(p, i)$ is satisfied and therefore $K_p^{i+1} \geq K_p^i$. We have P_i true.

- Inductive step ($0 < n \leq i$): Assume P_n and $K_p^{i+1} \neq \varnothing$ for a $p \in \mathbb{B}^{=n-1}$. By point 2 we have K_{p0}^{i+1} and K_{p1}^{i+1} different from $\varnothing$ and therefore, by inductive hypothesis that $K_{p1}^i, K_{p0}^i \neq \varnothing$ and also $K_{p1}^{i+1} \geq K_{p1}^i$ and $K_{p0}^{i+1} \geq K_{p0}^i$. Thus, by definition of a lower approximation, we obtain respectively $\phi_{lc}^*(p1, K_{p1}^{i+1}) \geq \phi_{lc}^*(p1, K_{p1}^i)$ and $\phi_{lc}^*(p0, K_{p0}^{i+1}) \geq \phi_{lc}^*(p0, K_{p0}^i)$. We deduce

  $$\phi_{lc}^*(p1, K_{p1}^{i+1}) + \phi_{lc}^*(p0, K_{p0}^{i+1}) \geq \phi_{lc}^*(p1, K_{p1}^i) + \phi_{lc}^*(p0, K_{p0}^i).$$

  Now by ♯ and knowing that $K_p^{i+1} \neq \varnothing$, we necessarily have that

  $$K_p^{i+1} = \min\{K \mid \phi_{lc}^*(p, K) \geq \phi_{lc}^*(p1, K_{p1}^{i+1}) + \phi_{lc}^*(p0, K_{p0}^{i+1})\}$$

  Thus, since $\phi_{lc}^*(p, \cdot)$ is non-decreasing, we have assignment (that is $K_p^i \neq \varnothing$) in ♯ of

  $$K_p^i = \min\{K \mid \phi_{lc}^*(p, K) \geq \phi_{lc}^*(p1, K_{p1}^i) + \phi_{lc}^*(p0, K_{p0}^i)\}$$

  such that $K_p^{i+1} \geq K_p^i$. This establishes the inductive step.

4) Assume $K_x^i \neq \varnothing$ and $x \in dom(\nu_i)$. Since $K_x^i \neq \varnothing$, by step 3, $\nu_i(x) = \phi_{lc}^*(x, K_x^i)$. The inequality $1 \geq \phi_{lc}^*(x, K_x^i)$ is guaranteed by the construction of K_x^i:

- If $|x| = i$, K_x^i (which equals i) is assigned in step 1 only if the condition $1 \geq \phi_{lc}^*(x, i)$ is satisfied. As $K_x^i \neq \varnothing$, this condition is true.
- If $|x| < i$, K_x^i is chosen according to ♯, which imposes $1 \geq \phi_{lc}^*(x, K)$ for the chosen K.

In both cases, $1 \geq \phi_{lc}^*(x, K_x^i)$, and therefore $1 \geq \nu_i(x)$. ■

Lemma: Partial mapping from ϕ^*_{lc} to ν_i

1. There exists an $r \in \mathcal{N} \cup \infty$ such that for all $i \in \{0, 1, 2, \ldots, r\}$ we have

$$\forall x \in \mathbb{B}^*, \quad x \in dom(\nu_i)$$

2. For all $i \in \{0, 1, 2, \ldots, r\}$, we have that ν_i are semimeasures satisfying

$$\forall x \in \mathbb{B}^*, \quad \nu_0(x) \leq \nu_1(x) \leq \ldots \leq \nu_r(x).$$

3. If $\phi^*_{lc} \xrightarrow{lc} \mu$ a semimeasure then $r = \infty$ and furthermore

$$\forall x \in \mathbb{B}^*, \quad \lim_{i \to \infty} \nu_i(x) = \mu(x).$$

Proof. 1) Let $x \in \mathbb{B}^*$. Let us define the set

$$S_x := \{i \in \mathcal{N} \mid x \notin dom(\nu_i)\}$$

By construction of ν_i we note that $x \notin dom(\nu_i)$ if and only if the 'Else' branch of step 1 is executed or if the search in $\sharp$ fails. This is equivalent to the existence of a string $p \in \mathbb{B}^{\leq i}$ for which K^i_p is not assigned, in other words,

$$i \in S_x \iff \exists p \in \mathbb{B}^{\leq i} \text{ such that } K^i_p = \varnothing.$$

Consider $i \in S_x$. Thus there exists $p \in \mathbb{B}^{\leq i}$ such that $K^i_p = \varnothing$. According to the contrapositive of point 3 of the previous Lemma (i.e. 'If $K^i_p = \varnothing$ then $K^{i+1}_p = \varnothing$'), we have $K^{i+1}_p = \varnothing$ and therefore $i + 1 \in S_x$. By repeating the reasoning starting from $K^{i+1}_p = \varnothing$ we show $K^{i+2}_p = \varnothing$ and so on.

Consequently, if $i \in S_x$, then $k \in S_x$ for all $k \geq i$. The set S_x is thus of the form $\{r, r + 1, r + 2, \ldots\}$ for $r = \min S_x$ or $r = \infty$ if it does not exist.

- If $r = \infty$ then $S_x = \emptyset$. We thus have $x \in dom(\nu_i)$ for all $i \in \mathcal{N}$. Point 1 of the lemma is satisfied.
- If $r \in \mathcal{N}$ then $S_x = \{r, r + 1, r + 2, \ldots\}$. Point 1 of the lemma is then satisfied:
  * If $i < r$, then $i \notin S_x$ and $x \in dom(\nu_i)$.
  * If $i \geq r$, then $i \in S_x$ and $x \notin dom(\nu_i)$.

2) Let $i \in \{0, 1, \ldots, r\}$ be an integer. Let us show first that ν_i is a semimeasure. Let x be in $\mathbb{B}^*$.

By $\sharp$ the variable K^i_ε is assigned if and only if $1 \geq \phi^*_{lc}(x, K^i_\varepsilon)$ hence $1 \geq \nu_i(\varepsilon)$. Moreover if $K^i_\varepsilon = \varnothing$ then by the 'Else' branch of step 3 we have $\nu_i(x) = 0$. It remains for us to verify $\nu_i(x) \geq \nu_i(x0) + \nu_i(x1)$. Three cases are possible:

- Case $|x| > i$: In particular $|x0| > i$ and $|x1| > i$. By point 1 of the previous lemma, $K^i_x = \varnothing$, $K^i_{x0} = \varnothing$ and $K^i_{x1} = \varnothing$. By the 'Else' branch of step 3 we have $\nu_i(x) = 0$, $\nu_i(x0) = 0$ and $\nu_i(x1) = 0$. The inequality $0 \geq 0 + 0$ is satisfied.
- Case $|x| = i$: Since $i \leq r$, we have $K^i_x \neq \varnothing$. The assignment is done in step 1, thus $K^i_x = i$. By point 4 of the previous lemma, $\mu_i(x) = \phi^*_{lc}(x, K^i_x) = \phi^*_{lc}(x, i)$. Since $|x0| = i + 1 > i$ and $|x1| = i + 1 > i$, by point 1 of the previous lemma, $K^i_{x0} = \varnothing$ and $K^i_{x1} = \varnothing$. By step 3, $\nu_i(x0) = 0$ and $\nu_i(x1) = 0$. The inequality becomes $\nu_i(x) \geq 0 + 0$. As ϕ^*_{lc} produces non-negative values, $\phi^*_{lc}(x, i) \geq 0$, thus the inequality is satisfied.

- Case $|x| < i$: Then $|x0| = |x| + 1 \leq i$ and $|x1| = |x| + 1 \leq i$. Since $i \leq r$, we have $K_x^i \neq \varnothing$ and by point 2 of the previous lemma $K_{x0}^i \neq \varnothing$ and $K_{x1}^i \neq \varnothing$. The assignment of K_x^i is done in step 2 via $\sharp$. This condition explicitly guarantees that K_x^i satisfies:

$$1 \geq \phi_{lc}^*(x, K_x^i) \geq \phi_{lc}^*(x0, K_{x0}^i) + \phi_{lc}^*(x1, K_{x1}^i).$$

  Since K_x^i, K_{x0}^i, K_{x1}^i are different from $\varnothing$, point 4 of the previous lemma applies and $\nu_i(p) = \phi_{lc}^*(p, K_p^i)$ for $p \in \{x, x0, x1\}$. By substituting into the previous inequality, we obtain $\nu_i(x) \geq \nu_i(x0) + \nu_i(x1)$.

It remains for us to show that $\nu_i(x) \leq \nu_{i+1}(x)$ for $i < r$. Let us now consider an integer i such that $i < r$. Then $i + 1 \leq r$. Consequently, for all $x \in \mathbb{B}^*$ we have $x \in dom(\nu_i)$ and $x \in dom(\nu_{i+1})$. Consider $x \in \mathbb{B}^*$.

- Case $|x| > i$: By point 1 of the previous lemma, $K_x^i = \varnothing$. By step 3, $\nu_i(x) = 0$. Since $\nu_{i+1}(x) \geq 0$ (as it is a semimeasure), we have $\nu_i(x) \leq \nu_{i+1}(x)$.

- Case $|x| \leq i$: Since $i \leq r$, we have $K_x^i \neq \varnothing$. As $i + 1 \leq r$, we also have $K_x^{i+1} \neq \varnothing$. By point 3 of the previous lemma, the hypothesis $K_x^{i+1} \neq \varnothing$ implies $K_x^{i+1} \geq K_x^i$. By definition of a strict lower approximation, the function $\phi_{lc}^*(x, \cdot)$ is non-decreasing. Thus $K_x^{i+1} \geq K_x^i$ implies:

$$\phi_{lc}^*(x, K_x^{i+1}) \geq \phi_{lc}^*(x, K_x^i)$$

  Since $K_x^i \neq \varnothing$ and $K_x^{i+1} \neq \varnothing$, and $x \in dom(\nu_{i+1})$ and $x \in dom(\nu_{i+1})$, point 4 of the previous lemma applies for i and $i + 1$: $\nu_i(x) = \phi_{lc}^*(x, K_x^i)$ and $\nu_{i+1}(x) = \phi_{lc}^*(x, K_x^{i+1})$. By combining these results, we obtain $\nu_{i+1}(x) \geq \nu_i(x)$.

3) Assume that there exists a semimeasure μ satisfying $\phi_{lc}^* \xrightarrow{lc} \mu$. First, let us show by contradiction that $r = \infty$. If $r < \infty$, the construction of ν_r fails for a string p of minimal length $|p| \leq r$.

- If $|p| = r$, the failure occurs at step $\dagger$, implying $1 < \phi_{lc}^*(p, r)$. Now, $\phi_{lc}^*(p, r) < \mu(p) \leq 1$ because ϕ_{lc}^* is a strict approximation and μ a semimeasure. This is a contradiction.

- If $|p| < r$, the failure occurs at step $\sharp$. This means that for all $K \geq r$, $\phi_{lc}^*(p, K) < \phi_{lc}^*(p0, K_{p0}^r) + \phi_{lc}^*(p1, K_{p1}^r)$. By passing to the limit $K \to \infty$, we obtain $\mu(p) \leq \phi_{lc}^*(p0, K_{p0}^r) + \phi_{lc}^*(p1, K_{p1}^r)$. By the strict character of the approximation, the right hand side is strictly less than $\mu(p0) + \mu(p1)$. We thus obtain $\mu(p) < \mu(p0) + \mu(p1)$, which contradicts μ being a semimeasure.

Both cases leading to a contradiction, we conclude that $r = \infty$. Consequently, the sequence $(\nu_i(x))_{i \in \mathcal{N}}$ is well-defined for all x. It is non-decreasing (point 2 of the lemma) and bounded by $\mu(x)$ because $\nu_i(x) = \phi_{lc}^*(x, K_x^i) < \mu(x)$. It therefore converges to a limit $L \leq \mu(x)$. By construction, $K_x^i \geq i$ for all $i \geq |x|$, which implies $\lim_{i \to \infty} K_x^i = \infty$. By composition of limits:

$$\lim_{i \to \infty} \nu_i(x) = \lim_{i \to \infty} \phi_{lc}^*(x, K_x^i) = \mu(x)$$

■

Lemma: Mapping $\langle \phi^*_{lc} \rangle \xrightarrow{\langle \mathcal{T}^<_{lc} \rangle \Rightarrow \langle \mathcal{T}_{sc} \rangle} \langle \phi_{sc} \rangle$

Let ϕ^*_{lc} be a strict lower approximation. Then there exists a function ϕ_{sc} such that:

1. ϕ_{sc} is a lower approximation of a semimeasure.
2. If ϕ^*_{lc} is a strict lower approximation of a semimeasure μ, then ϕ_{sc} is a lower approximation of μ.

Proof. 1) We reuse the definitions and notations of the previous lemma, that is, there exist partial computable functions $(\nu_i)_{i\in\mathcal{N}}$ and an integer $r \in \mathcal{N} \cup \{\infty\}$ which satisfy the points of the lemma. Let us define the function $\mu' : \mathbb{B}^* \to \mathbb{R} \cup \{\infty\}$ as:

$$\mu(x) := \sup_{0\leq i<r} \nu_i(x)$$

The function μ' is a semimeasure. Indeed, the semimeasure properties of the ν_i (namely $\nu_i(\varepsilon) \leq 1$ and $\nu_i(x) \geq \nu_i(x0) + \nu_i(x1)$) are preserved by taking the supremum. Let us introduce the notation $\nu_i(x)[k]$ corresponding to the result of the computation of $\nu_i(x)$ if it halts in at most k steps. Let us set, for all $x \in \mathbb{B}^*$ and $k \in \mathcal{N}$:

$$\phi_{sc}(x, k) := \max\left(\{0\} \cup \{\nu_i(x)[k] \mid 0 \leq i \leq k \text{ and } \nu_i(x)[k] \text{ is defined}\}\right)$$

Let us verify that ϕ_{sc} is a lower approximation of μ'.

Let us justify that ϕ_{sc} is total computable. For fixed inputs x, k and i, the computation $\nu_i(x)[k]$ requires a finite number of steps. Thus, knowing that $0 \leq i \leq k$, the number of steps to calculate $\phi_{sc}(x, k)$ is therefore finite.

Let us justify that $\phi_{sc}(x, \cdot)$ is non-decreasing for all x. The function ϕ_{sc} seeks the maximum of a set that grows in the sense of inclusion with respect to k. Consequently, $\phi_{sc}(x, k) \leq \phi_{sc}(x, k+1)$ for all k.

It remains for us to prove that $\lim_{k\to\infty} \phi_{sc}(x, k) = \mu'(x)$. Let $x \in \mathbb{B}^*$. By construction, each $\nu_i(x)[k]$ is less than or equal to $\nu_i(x)$, which is itself less than or equal to $\mu'(x)$. Thus, $\phi_{sc}(x, k) \leq \mu'(x)$ for all k, and the limit is bounded by $\mu'(x)$. Conversely, for all $y < \mu'(x)$, there exists by definition of the supremum an $i_0 < r$ such that $\nu_{i_0}(x) > y$. The computation of $\nu_{i_0}(x)$ being a terminating procedure, there exists a number of steps k_0 after which it produces this result. Then, for all $k \geq \max(i_0, k_0)$, the term $\nu_{i_0}(x)[k]$ is defined and equal to $\nu_{i_0}(x)$. Therefore, for these k, we have $\phi_{sc}(x, k) \geq \nu_{i_0}(x) > y$. Since this is true for all $y < \mu'(x)$, the limit is equal to $\mu'(x)$.

2) Assume that ϕ^*_{lc} is a strict lower approximation of a semimeasure μ. According to point 3 of Lemma 2, this hypothesis implies that $r = \infty$ and that for all $x \in \mathbb{B}^*$, the sequence $(\nu_i(x))$ converges to $\mu(x)$. In this case, the semimeasure μ' constructed above is μ:

$$\mu'(x) = \sup_{i\geq 0} \nu_i(x) = \lim_{i\to\infty} \nu_i(x) = \mu(x)$$

Since we have established that ϕ_{sc} is a lower approximation of μ_{sm}, it is consequently a lower approximation of μ. ■

Definition-Theorem: Compact enumeration of lower approximations of Π_{sm}

We call a compact enumeration of lower approximations of semimeasures a total computable function $\pi_{sm} : i \in \mathcal{N} \mapsto \langle\, \phi_{sm,i} \,\rangle \in \langle\, \mathcal{T}_{sc} \,\rangle$ such that

$$\forall \mu \in \Pi_{sm}, \quad \exists i \in \mathcal{N}, \quad \pi_{sm}(i) = \langle\, \phi_{sm,i} \,\rangle \text{ and } \phi_{sm,i} \xrightarrow{lc} \mu$$

with the convention that $\langle\, \phi_{sm,i} \,\rangle$ is the code of a Turing machine in $\langle\, \mathcal{T}_{sc} \,\rangle$ that computes the function $\phi_{sm,i}$.

Proof. Let $\pi_{lc}^{<}$ be an effective enumeration of $\langle\, \mathcal{T}_{lc}^{<} \,\rangle$. Let us set π_{sm} a computable function with the following behavior:

For an input $i \in \mathcal{N}$:

- Assign $\langle\, \phi_{lc,i}^{*} \,\rangle \leftarrow \pi_{lc}^{<}(i)$.
- Apply the Mapping $\langle\, \phi_{lc,i}^{*} \,\rangle \xrightarrow{\langle \mathcal{T} \rangle \Rightarrow \langle \mathcal{T}_{sc} \rangle} \langle\, \phi_{sm,i} \,\rangle$.
- Accept the input with $\langle\, \phi_{sm,i} \,\rangle$ as output.

Let us prove that $\pi_{lc}^{<}$ is a compact enumeration of $\langle\, \mathcal{T}_{lc}^{<} \,\rangle$. Let ϕ_{lc} be a lower approximation of a function f. By definition of π, there exists $i \in \mathcal{N}$ such that $\pi_{lc}^{<}(i) = \langle\, \phi_{lc}^{*} \,\rangle$. Thus according to the Mapping, $\langle\, \phi_{sm,i} \,\rangle$ is the code of a lower approximation which satisfies $\phi_{sm,i} \xrightarrow{l.c} \mu$. ■

1.1.3 Universal lower semicomputable semimeasure

Definition-Theorem: Universal semimeasure

A function $\mathbf{M} : \mathbb{B}^* \to \mathbb{R} \cup \{\infty\}$ is a universal semimeasure if it satisfies the following two conditions:

1. $\mathbf{M}$ is a lower semicomputable semimeasure (i.e., $\mathbf{M} \in \Pi_{sm}$).
2. For every semimeasure $\mu \in \Pi_{sm}$, there exists a constant $c_\mu > 0$ such that for all $x \in \mathbb{B}^*$:
$$\mathbf{M}(x) \geq c_\mu \cdot \mu(x)$$

Moreover, such a universal semimeasure exists.

Proof. Let $\pi : i \mapsto \langle\, \phi_{sm,i} \,\rangle$ be a compact enumeration of lower approximations. Let $\phi_{sm,i} \xrightarrow{lc} \mu_i$ for every integer i. We then have that $(\phi_{sm,i})_{i \in \mathcal{N}}$ is an enumeration of all lower semicomputable semimeasures. Let $\alpha : \mathcal{N} \to \mathbb{R}^+$ be a function in $\Pi_{\nearrow}$ lower approximated by ϕ_α and satisfying $\alpha(i) > 0$ for all i and $\sum_{i=0}^{\infty} \alpha(i) \leq 1$. Such a function exists, for example the computable function $\alpha(i) = 2^{-(i+1)}$. Let us define the function $\mathbf{M}$ by:

$$\mathbf{M}(x) := \sum_{i=0}^{\infty} \alpha(i) \mu_i(x)$$

Let us show that $\mathbf{M}$ is a lower semicomputable semimeasure and then that it is universal:

Let us justify that $\mathbf{M} \in \Pi_{sm}$. For this, let us set the function $\Phi(x,k) := \sum_{i=0}^{k} \phi_\alpha(i,k)\phi_{sm,i}(x,k)$ and show that it is a total lower approximation of $\mathbf{M}$. We already have, for any string x, that $\Phi(x,\cdot)$ is non-decreasing by monotonicity of ϕ_α and $\phi_{sm,i}(x,\cdot)$. We obtain by letting $k \to \infty$:

- $\mathbf{M}(\varepsilon) = \sum_i \alpha(i)\mu_i(\varepsilon) \leq \sum_i \alpha(i) \leq 1$.
- $\mathbf{M}(x0) + \mathbf{M}(x1) = \sum_i \alpha(i)(\mu_i(x0) + \mu_i(x1)) \leq \sum_i \alpha(i)\mu_i(x) = \mathbf{M}(x)$.

Let us justify that $\mathbf{M}$ is universal. For any semimeasure $\mu_j \in \Pi_{sm}$ and any string $x \in \mathbb{B}^*$,

$$\mathbf{M}(x) = \sum_{i=0}^{\infty} \alpha(i)\mu_i(x) \geq \alpha(j)\mu_j(x)$$

Let us then set $c_j = \alpha(j)$, we obtain $\mathbf{M}(x) \geq c_j \cdot \mu_j(x)$. ■

Convention:

In the rest of this course, we set $\mathbf{M}$ to be any universal lower semicomputable semimeasure.

1.2 Solomonoff Semimeasure

1.2.1 Definitions

Definition-Property: Solomonoff Semimeasure

Let M_{mt} be a monotone Turing machine. The Solomonoff semimeasure associated with M_{mt}, denoted by $\lambda_{M_{mt}}$, is the function from $\mathbb{B}^*$ to $[0, 1]$ defined by:

$$\lambda_{M_{mt}}(x) := \sum_{\lfloor p \,|\, M_{mt}(p)=x* \rfloor} 2^{-\ell(p)}$$

This function is a lower semicomputable semimeasure. The set of semimeasures thus defined is denoted by Π_{Sol}.

Proof. Let M_{mt} be a monotone Turing machine. For any string $y \in \mathbb{B}^*$, we denote $S_y = \{q \in \mathbb{B}^* \mid M_{mt}(q) = y*\}$ and $P_y = \lfloor S_y \rfloor$ the set of minimal programs in S_y for the prefix order.

The first condition, $\lambda_{M_{mt}}(\epsilon) \leq 1$, is a direct consequence of Kraft's inequality. The set P_ϵ is, by definition of $\lfloor \cdot \rfloor$, a prefix-free set. Thus, $\sum_{p \in P_\epsilon} 2^{-\ell(p)} \leq 1$.

It remains to demonstrate the second condition, $\lambda_{M_{mt}}(x) \geq \lambda_{M_{mt}}(x0) + \lambda_{M_{mt}}(x1)$ for all $x \in \mathbb{B}^*$. We first establish that the union $P_{x0} \cup P_{x1}$ is a prefix-free set. The sets P_{x0} and P_{x1} are disjoint, because the output of a single program cannot simultaneously start with the incompatible prefixes $x0$ and $x1$. Moreover, no element $p_1 \in P_{x0}$ can be a prefix of an element $p_2 \in P_{x1}$ (or conversely). Indeed, if $p_1 \leq_p p_2$, the monotonicity of M_{mt} would imply $M_{mt}(p_1) \leq_p M_{mt}(p_2)$, which would contradict the incompatibility of the output prefixes $x0$ and $x1$. Since P_{x0} and P_{x1} are themselves prefix-free, their union is also so.

Let us next show that every string in $P_{x0} \cup P_{x1}$ has a prefix in P_x. Let $p \in P_{x0} \cup P_{x1}$. By definition, $M_{mt}(p)$ starts with $x0$ or $x1$, therefore $p \in S_x$. The set $\{t \in S_x \mid t \leq_p p\}$ is non-empty (since it contains p) and finite. Let p' be one of its minimal elements for the order $\leq_p$. By construction, $p' \in S_x$ and $p' \leq_p p$. Moreover, p' is necessarily an element of P_x. Suppose, for the sake of contradiction, that $p' \notin P_x$. Then, by definition of P_x, there would exist a proper prefix $s \leq_p p'$ such that $s \in S_x$. By transitivity, $s \leq_p p$, which would contradict the minimality of p' in the considered set. Therefore, $p' \in P_x$.

Let us show that $\bigcup_{p \in P_{x0} \cup P_{x1}} \Gamma_p \subseteq \bigcup_{p \in P_x} \Gamma_p$. Let Γ_p be a cylinder in the left union, with $p \in P_{x0} \cup P_{x1}$. According to what we have shown in the previous paragraph, there exists a $p' \in P_x$ such that $p' \leq_p p$. We then deduce the inclusion of the cylinders $\Gamma_p \subseteq \Gamma_{p'}$. As $p' \in P_x$, the cylinder $\Gamma_{p'}$ is itself included in the right union. Consequently, Γ_p is also included in the right union, which proves the global inclusion.

Since P_x and $P_{x0} \cup P_{x1}$ are prefix-free sets, the cylinders in each union are disjoint. The value of $\lambda_{M_{mt}}$ thus coincides with the measure of the corresponding union. We have:

$$\lambda_{M_{mt}}(x) = \sum_{p' \in P_x} 2^{-\ell(p')} = \lambda \left(\bigcup_{p' \in P_x} \Gamma_{p'} \right)$$

and, using that P_{x0} and P_{x1} are disjoint,

$$\lambda_{M_{mt}}(x0) + \lambda_{M_{mt}}(x1) = \sum_{p \in P_{x0} \cup P_{x1}} 2^{-\ell(p)} = \lambda \left(\bigcup_{p \in P_{x0} \cup P_{x1}} \Gamma_p \right)$$

The monotonicity of the measure λ, applied to the inclusion of the unions of cylinders, then leads to the sought inequality:

$$\lambda_{M_{mt}}(x0) + \lambda_{M_{mt}}(x1) \leq \lambda_{M_{mt}}(x)$$

This completes the proof that $\lambda_{M_{mt}}$ is a semimeasure. ■

1.2.2 $\Pi_{Sol} \subset \Pi_{sm}$

Lemma:

Let $(L_k)_{k \in \mathcal{N}}$ be a sequence of languages in $\mathbb{B}^*$ increasing in the sense of inclusion. Let us set $L = \bigcup_{k \in \mathcal{N}} L_k$ and $\Lambda_k = \bigcup_{p \in \lfloor L_k \rfloor} \Gamma_p$. Then:

1. The sequence of sets $(\Lambda_k)_{k \in \mathcal{N}}$ is increasing for inclusion.
2. The limit of this sequence is $\lim_{k \to \infty} \Lambda_k = \bigcup_{p \in \lfloor L \rfloor} \Gamma_p$.

Proof. 1) Let $w \in U_k$. By definition, there exists $p \in \lfloor L_k \rfloor$ such that $p \leq_p w$. Since $L_k \subseteq L_{k+1}$, we have $p \in L_{k+1}$. There exists therefore a minimal prefix $q \in \lfloor L_{k+1} \rfloor$ of p, such that $q \leq_p p$. By transitivity, $q \leq_p w$, thus $w \in \Gamma_q \subseteq \Lambda_{k+1}$. The inclusion $\Lambda_k \subseteq \Lambda_{k+1}$ is thus demonstrated.

2) The sequence (Λ_k) being increasing, its limit is its union. The equality to prove becomes:

$$\bigcup_{k \in \mathcal{N}} \Lambda_k = \bigcup_{p \in \lfloor L \rfloor} \Gamma_p$$

We demonstrate it by double inclusion:

$\subset$: Let $w \in \bigcup_k U_k$. There exists k and $p \in \lfloor L_k \rfloor$ such that $p \leq_p w$. Since $p \in L_k \subseteq L$, there exists a minimal prefix $q \in \lfloor L \rfloor$ such that $q \leq_p p$. By transitivity, $q \leq_p w$, which implies $w \in \bigcup_{p' \in \lfloor L \rfloor} \Gamma_{p'}$.

$\supset$: Let $w \in \bigcup_{p \in \lfloor L \rfloor} \Gamma_p$. There exists $p \in \lfloor L \rfloor$ such that $p \leq_p w$. By definition of $\lfloor L \rfloor$, we have $p \in L$ and no proper prefix of p belongs to L. As $L = \bigcup_k L_k$, there exists a rank k_0 such that $p \in L_{k_0}$. Moreover, no proper prefix of p belonging to L, consequently none belongs to L_{k_0}. These two points are the very definition of $p \in \lfloor L_{k_0} \rfloor$. Thus, $w \in \Gamma_p \subseteq \Lambda_{k_0} \subseteq \bigcup_k \Lambda_k$. ■

Theorem: $\Pi_{Sol} \subseteq \Pi_{sm}$

For any monotone Turing machine M_{mt}, the Solomonoff semimeasure $\lambda_{M_{mt}}$ is lower semicomputable.

Proof. Let M_{mt} be a monotone Turing machine. Let ϕ_{sc} operate as follows:

For an input $\langle\, x, k \,\rangle$ in $\langle\, \mathbb{B}^*, \mathcal{N} \,\rangle$ do:

- For $p_0, p_1, \ldots$ in lexicographical order, assign

$$P_k \leftarrow \{p_i \mid x \leq_p M_{mt}(p_i \mid k) \text{ and } 0 \leq i \leq k\}$$

- Calculate $\lfloor P_k \rfloor$ then write $\sum_{p \in \lfloor P_k \rfloor} 2^{-\ell(p)}$ as output before accepting.

Let us justify that ϕ_{sc} is total computable. Consider the call $\phi_{sc}(x, k)$. The assignment of P_k requires the computation of $M_{mt}(p_i \mid k)$ for $0 \leq i \leq k$ which requires a finite number of steps. The computation of $\lfloor P_k \rfloor$ is effective: Keep the first element of P_k only if all other elements of P_k are not compatible with the first. Do the same with the second, the third ... until the last. The set of kept elements is then $\lfloor P_k \rfloor$. Finally, the output $\sum_{p \in P'_k} 2^{-\ell(p)}$ is calculated before acceptance. The call $\phi_{sc}(x, k)$ terminates.

Let $x \in \mathbb{B}^*$. Let us denote $P_{x,k}$ the sets assigned in the calls $\phi_{sc}(x, k)$ for $k \in \mathcal{N}$. We note that $(P_k)_{k \in \mathcal{N}}$ is increasing by inclusion, and its union $P := \bigcup_k P_k$ is precisely the set $\{p \in \mathbb{B}^* \mid M_{mt}(p) = x*\}$. This will allow us to apply the previous lemma in what follows.

The output of the machine is $\phi_{sc}(x, k) = \sum_{p \in \lfloor P_k \rfloor} 2^{-\ell(p)}$. By definition:

$$\phi_{sc}(x, k) = \lambda \left(\bigcup_{p \in \lfloor P_k \rfloor} \Gamma_p \right)$$

Moreover, the Solomonoff semimeasure is written $\lambda_{M_{mt}}(x) = \lambda(\bigcup_{p \in \lfloor L_x \rfloor} \Gamma_p)$. Let us show in the following that ϕ_{sc} is a lower approximation of $\lambda_{M_{mt}}$:

- Let us prove that $\phi_{sc}(x, \cdot)$ is non-decreasing. According to point 1 of the previous lemma, the sequence of sets $\Lambda_k := \bigcup_{p \in \lfloor P_k \rfloor} \Gamma_p$ is increasing for inclusion. By monotonicity of the measure λ, the sequence of their measures $(\lambda(\Lambda_k))_{k \in \mathcal{N}}$ is non-decreasing. Now, $\phi_{sc}(x, k) = \lambda(\Lambda_k)$, therefore $\phi_{sc}(x, k) \leq \phi_{sc}(x, k+1)$.

- Let us prove that $\lim_{k \to \infty} \phi_{sc}(x, k) = \lambda_{M_{mt}}(x)$. By continuity of the measure on an increasing sequence of sets, the limit of the measure is the measure of the limit:

$$\lim_{k \to \infty} \phi_{sc}(x, k) = \lim_{k \to \infty} \lambda(\Lambda_k) = \lambda \left(\lim_{k \to \infty} \Lambda_k \right)$$

  Using point 2 of the previous lemma, we deduce:

$$\lim_{k \to \infty} \Lambda_k = \bigcup_{p \in \lfloor P \rfloor} \Gamma_p$$

Combining these two equalities, we obtain:

$$\lim_{k\to\infty} \phi_{sc}(x,k) = \lambda\left(\bigcup_{p\in\lfloor P\rfloor} \Gamma_p\right) = \lambda_{M_{mt}}(x)$$

In summary ϕ_{sc} is total, non-decreasing with respect to its second component, and converges to $\lambda_{M_{mt}}(x)$. It is therefore a lower approximation, which proves that $\lambda_{M_{mt}} \in \Pi_{sm}$. ■

1.2.3 $\Pi_{Sol} \supset \Pi_{sm}$

Lemma: Construction of dyadic interval Θ_k

Let $\phi_{lc}^{\sim}$ be a compatible increasing dyadic approximator. Let $\Delta_k(x) := \phi_{lc}^{\sim}(x,k) - \phi_{lc}^{\sim}(x,k-1)$, with the convention $\Delta_{-1}(x) = 0$. For all $k \in \mathcal{N}$ and $x = x_1x_2\dots x_l \in \mathbb{B}^*$, by recurrence on $0 \le i < l$, let us also set

$$\nu_k(x_{1:i+1}) := \begin{cases} \nu_k(x_{1:i}) & \text{if } x_{i+1} = 0 \\ \nu_k(x_{1:i}) + \Delta_k(x_{1:i}0) & \text{if } x_{i+1} = 1 \end{cases}$$

where $\nu_k(\varepsilon) := \sum_{j=0}^{k-1} \Delta_j(\varepsilon)$ and $x_{1:0} := \varepsilon$. We then define the sets:

$$\Theta_k(x) := [\nu_k(x), \nu_k(x) + \Delta_k(x)[\quad \text{and} \quad \Theta(x) := \bigcup_{k=0}^{\infty} \Theta_k(x)$$

We then have for all $x, y \in \mathbb{B}^*$ and $k \in \mathcal{N}$ that

1. $x \le_p y \implies \Theta_k(y) \subseteq \Theta_k(x)$
2. $\forall k' \neq k, \quad \Theta_k(x) \cap \Theta_{k'}(y) = \emptyset$
3. x, y incompatible $\implies \Theta_k(x) \cap \Theta_k(y) = \emptyset$
4. $\lambda(\Theta_k(x)) = \Delta_k(x)$
5. $\nu_k(x)$ is dyadic and $\Theta_k(x)$ has dyadic endpoints (i.e. of the form $[\frac{n}{2^i}, \frac{n'}{2^{i'}}[$)

Proof. Let $x, y \in \mathbb{B}^*$ and $k \in \mathcal{N}$. Let us show the points of the lemma in order.

1) Assume $x <_p y$ and let us show that $\Theta_k(y) \subsetneq \Theta_k(x)$. Let $\sigma \in \mathbb{B}^* \setminus \varepsilon$ such that $y = x\sigma$. We proceed by induction on the length of σ:

Base case ($|\sigma| = 1$): Let $y = x\sigma$ with $|\sigma| = 1$. We thus have $\sigma \in \mathbb{B}$, that is to say two possible cases:

- If $\sigma = 0$ then $y = x0$. Exploiting the definition $\nu_k(x0) = \nu(x)$ we obtain

$$\Theta_k(x0) = [\nu_k(x0), \nu_k(x0) + \Delta_k(x0)[= [\nu_k(x), \nu_k(x) + \Delta_k(x0)[$$

By points 3 and 6 of the definition of a compatible increasing dyadic approximation machine we have respectively $\Delta_k(x1) \ge 0$ and $\Delta_k(x) \ge \Delta_k(x0) + \Delta_k(x1)$ which gives

$$\nu_k(x) + \Delta_k(x0) \le \nu_k(x) + \Delta_k(x0) + \Delta_k(x1) \le \nu_k(x) + \Delta_k(x)$$

Thus we obtain

$$\Theta_k(x0) = [\nu_k(x), \nu_k(x) + \Delta_k(x0)[\subset [\nu_k(x), \nu_k(x) + \Delta_k(x)[= \Theta_k(x)$$

that is $\Theta_k(x0) \subseteq \Theta_k(x)$.

- If $\sigma = 1$ then $y = x1$. Exploiting that, by definition, $\nu_k(x0) = \nu(x) + \Delta_k(x0)$ we obtain

$$\Theta_k(x1) = [\,\nu_k(x1), \nu_k(x1) + \Delta_k(x1)\,[= [\nu_k(x) + \Delta_k(x0), \nu_k(x) + \Delta_k(x0) + \Delta_k(x1)\,[$$

However, by points 3 and 6 of the definition of a compatible increasing dyadic approximation machine we have respectively $\Delta_k(x0) \geq 0$ and $\Delta_k(x) \geq \Delta_k(x0) + \Delta_k(x1)$ which gives

$$\begin{cases} \nu_k(x) + \Delta_k(x0) \geq \nu_k(x) \\ \nu_k(x) + \Delta_k(x0) + \Delta_k(x1) \leq \nu_k(x) + \Delta_k(x) \end{cases}$$

that is $\Theta_k(x1) \subset \Theta_k(x)$.

Inductive step ($|\sigma| = n$): Assume that for all σ of length $n \geq 1$ we have $\Theta_k(x\sigma) \subset \Theta_k(x)$. Let $y = x\sigma b$ with $|\sigma| = n$ and $b \in \{0, 1\}$. According to the base case, $\Theta_k((x\sigma)b) \subset \Theta_k(x\sigma)$. Moreover, by the induction hypothesis we have $\Theta_k(x\sigma) \subset \Theta_k(x)$. Which, by transitivity of inclusion, gives $\Theta_k(y) \subset \Theta_k(x)$ and thus proves the inductive step.

2) According to point 1, for all $x \in \mathbb{B}^*$, we have $\Theta_k(x) \subseteq \Theta_k(\varepsilon)$. It therefore suffices to show that the intervals $\Theta_k(\varepsilon)$ and $\Theta_{k'}(\varepsilon)$ are disjoint for $k \neq k'$.

Let us set for this $S_k := \sum_{j=0}^{k-1} \Delta_j(\varepsilon)$. By definition, $\Theta_k(\varepsilon) = [\nu_k(\varepsilon), \nu_k(\varepsilon) + \Delta_k(\varepsilon)] = [S_k, S_k + \Delta_k(\varepsilon)]$. The interval can thus be written $\Theta_k(\varepsilon) = [S_k, S_{k+1}]$.

Let k, k' in $\mathcal{N}$ be distinct. Without loss of generality, assume $k < k'$. We then have $k + 1 \leq k'$. The supremum of the interval for k is $\sup(\Theta_k(\varepsilon)) = S_{k+1}$. The infimum of the interval for k' is $\inf(\Theta_{k'}(\varepsilon)) = S_{k'}$.

By definition of S_k, and since $\Delta_j(\varepsilon) \geq 0$ for all j, the sequence $(S_k)_{k\in\mathcal{N}}$ is increasing. Therefore, $S_{k+1} \leq S_{k'}$. Thus, $\sup(\Theta_k(x)) \leq S_{k+1} \leq S_{k'} \leq \inf(\Theta_{k'}(y))$. The intervals $\Theta_k(x)$ and $\Theta_{k'}(y)$ are therefore disjoint.

3) Assume x, y are incompatible. Let z be their longest common prefix, that is $x = z0\sigma$ and $y = z1\tau$ (without loss of generality). By definition

$$\begin{cases} \Theta_k(z0) = [\nu_k(z0), \nu_k(z0) + \Delta_k(z0)[= [\nu_k(z), \nu_k(z) + \Delta_k(z0)[\\ \Theta_k(z1) = [\nu_k(z1), \nu_k(z1) + \Delta_k(z1)[= [\nu_k(z) + \Delta_k(z0), \nu_k(z) + \Delta_k(z0) + \Delta_k(z1)[\end{cases}$$

As $\Delta_k(z0) \geq 0$ we remark $\Theta_k(z0) \cap \Theta_k(z1) = \emptyset$. Moreover by point 1, knowing that $z0 \leq_p x$ and $z1 \leq_p y$, we have respectively $\Theta_k(x) \subset \Theta_k(z1)$ and $\Theta_k(y) \subset \Theta_k(z1)$. We deduce $\Theta_k(x) \cap \Theta_k(y) = \emptyset$.

4) Let $x \in \mathbb{B}^*$. By definition of $\Theta_k(x)$ we immediately obtain the result

$$\lambda(\Theta_k(x)) = (\nu_k(x) + \Delta_k(x)) - \nu_k(x) = \Delta_k(x)$$

5) Let us show that for all z, $\nu_k(z)$ is dyadic and $\Theta_k(z)$ is a dyadic interval. Let us proceed by induction on the size of $|z|$.

Base case ($|z| = 0$): Of course, $|z| = 0$ implies that $z = \varepsilon$. Now, dyadic numbers are closed under addition and subtraction. Thus $\nu_k(z) = \nu_k(\varepsilon) := \sum_{j=0}^{k-1} \Delta_j(\varepsilon)$ is dyadic and $\Theta_k(z) = [\nu_k(z), \nu_k(z) + \Delta_k(z)[$ is an interval with dyadic endpoints.

Inductive step ($|z| = n$): Assume that $\nu_k(z)$ is dyadic. By closure of dyadic numbers under addition, $\nu_k(z0) := \nu_k(z)$ and $\nu_k(z1) := \nu_k(z) + \Delta_k(z0)$ are dyadic. Thus we have [that] $\Theta(z0) = [\nu_k(z0), \nu(z0) + \Delta_k(z0)[$ and $\Theta(z1) = [\nu_k(z1), \nu(z1) + \Delta_k(z1)[$ are intervals with dyadic endpoints, which establishes the inductive step. ■

Algorithm: Mapping $\phi_{lc}^{\sim} \xrightarrow{\Pi_{sm} \Rightarrow \Pi_{Sol}} M_{mt}$

For $\phi_{lc}^{\sim}$ a compatible increasing dyadic lower approximator. Recall that I_p corresponds to the binary interval associated with $p \in \mathbb{B}^*$. Let us define the Turing machine M_{mt} which operates as follows:

Input : $p \in \mathbb{B}^*$, (and let $p := p_1 p_2 \ldots p_l$)

Step 1. Assign sequentially:

- $k_p \leftarrow \min\{j \in \mathcal{N} \mid I_p \subset [\sum_{t=0}^{j-1} \Delta(\varepsilon, t), \sum_{t=0}^{j} \Delta(\varepsilon, t)[\,\}$
  (Obtained by iterating $j = 0, 1, 2, \ldots$ up to the first j satisfying it, denoted by k_p)
- $\kappa(\varepsilon) \leftarrow \phi_{lc}^{\sim}(\varepsilon, k_p)$
- $i \leftarrow 0$
- $s_0 \leftarrow \varepsilon$

Step 2. Three cases are possible:

(i) Case $I_p \subset [\kappa(s_i), \kappa(s_i) + \Delta(s_i 0, k_p)\,[$ then assign:
  * $\kappa(s_i 0) \leftarrow \kappa(s_i)$
  * $s_{i+1} \leftarrow s_i 0$

(ii) Case $I_p \subset [\kappa(s_i) + \Delta(s_i 0, k_p), \kappa(s_i) + \Delta(s_i 1, k_p) + \Delta(s_i 0, k_p)\,[$ then assign:
  * $\kappa(s_i 1) \leftarrow \Delta(s_i 0, k_p) + \kappa(s_i)$
  * $s_{i+1} \leftarrow s_i 1$

(iii) If cases (i) and (ii) fail then:
  * Loop forever $\circlearrowleft M_{mt}(p)$

Write s_{i+1} to output (Appending to the end of the output tape to preserve the monotonicity of the output), assign $i \leftarrow i + 1$ then Repeat Step 2.

Note that we cannot use the fact that M_{mt} is a monotone Turing machine in the following lemma. This is shown in the subsequent lemma.

Lemma: M_{mt} corresponds to Θ_k

Let us resume the definitions of $\nu_k(x)$ and $\Theta_k(x)$ from the previous lemma. Let M_{mt} be the Turing machine defined by the algorithm below for a $\phi_{lc}^{\sim}$. For $k \in \mathcal{N}$ and $x \in \mathbb{B}^*$ let us set the set

$$P_x^k := \{p \in \mathbb{B}^* \mid M_{mt}(p) = x* \ \text{ and } \ I_p \subset \Theta_k(x)\}$$

We then have for all $k \in \mathcal{N}$ and $x \in \mathbb{B}^*$ that:

1. For all p such that $I_p \subset \Theta_k(x)$, we have $k_p = k$.
2. For all p such that $I_p \subset \Theta_k(x)$, we have $M_{mt}(p) = x*$.
3. $\bigcup_{p \in P_x^k} I_p = \Theta_k(x)$.

Proof. Let $k \in \mathcal{N}$ and $x = x_1 \dots x_l \in \mathbb{B}^*$.

1) Assume $I_p \subset \Theta_k(x)$. By step 1 of the algorithm, k_p is defined as the smallest integer j such that $I_p \subset \left[\sum_{t=0}^{j-1} \Delta(\varepsilon, t), \sum_{t=0}^{j} \Delta(\varepsilon, t)\right[$. By definition of the intervals Θ, this condition is $I_p \subset \Theta_{k_p}(\varepsilon)$.

On the other hand, our initial hypothesis is $I_p \subset \Theta_k(x)$. Since $\varepsilon \leq_p x$, point 1 of the lemma "Construction of dyadic interval" guarantees that $\Theta_k(x) \subseteq \Theta_k(\varepsilon)$. By transitivity of inclusion, we therefore have $I_p \subset \Theta_k(\varepsilon)$.

We have thus shown that $I_p \subset \Theta_{k_p}(\varepsilon)$ and $I_p \subset \Theta_k(\varepsilon)$. Now, point 2 of the lemma "Construction of dyadic interval" states that the sets $\Theta_j(\varepsilon)$ are disjoint for $j \neq j'$. Since I_p is a non-empty interval contained in their intersection, it is necessary that the indices be equal. Thus, $k_p = k$.

2) Let $p \in \mathbb{B}^*$ and assume $I_p \subset \Theta_k(x)$. Let $(s_i)_i$ and $\kappa(\cdot)$ be the variables assigned in the call $M_{mt}(p)$. We will show by induction on $i \in \{0, \dots, l\}$ that the algorithm successively generates the prefixes of x. Let us set the loop invariant $\mathcal{I}(i)$, which is verified at the beginning of iteration $i+1$ of step 2:

$$\mathcal{I}(i): \quad s_i = x_{1:i}, \quad \kappa(s_i) = \nu_k(s_i), \quad \text{and} \quad I_p \subset \Theta_k(s_i)$$

Initialization ($i = 0$): By step 1 of M_{mt} we necessarily have

- $s_0 = \varepsilon = x_{1:0}$
- $k_p = k$ (according to point 1). Thus $\kappa(\varepsilon) = \phi_{lc}^{\sim}(\varepsilon, k) = \sum_{j=0}^{k-1} \Delta_j(\varepsilon) = \nu_k(\varepsilon) = \nu_k(s_0)$
- The hypothesis $I_p \subset \Theta_k(x)$ and the fact that $x_{1:0} = \varepsilon \leq_p x$ imply $I_p \subset \Theta_k(x) \subseteq \Theta_k(x_{1:0}) = \Theta_k(s_0)$

The invariant $I(0)$ is therefore verified.

Inductive step ($0 \leq i < l$): Assume the invariant $\mathcal{I}(i)$ is true, that is $s_i = x_{1:i}$, $\kappa(s_i) = \nu_k(s_i)$ and $I_p \subset \Theta_k(s_i)$. Since $i < l$, the algorithm must necessarily still write bits to the output, hence the exclusion of case (iii), (because this case implies that ↺ M_{mt} loops forever without writing any more to the output). We then have the execution of either case (i) or case (ii) to produce the next bit, x_{i+1}:

- Case $x_{i+1} = 0$: By the invariant $I(i)$ we have $I_p \subset \Theta_k(x_{1:i})$. Moreover, by the inclusion $\Theta_k(x) \subseteq \Theta_k(x_{1:i+1})$ (by point 1 of the lemma "Construction of dyadic interval"), this gives $I_p \subset \Theta_k(x_{1:i+1})$.

By definition of $\Theta_k(x_{1:i}0)$ and using $\mathcal{I}(i)$ and $\nu_k(x_{1:i}0) = \nu_k(x_{1:i})$, we have:

$$\Theta_k(x_{1:i}0) = [\nu_k(x_{1:i}), \nu_k(x_{1:i}) + \Delta_k(x_{1:i}0)[= [\kappa(s_i), \kappa(s_i) + \Delta_k(s_i0)[$$

This is precisely condition (i), hence its execution:

$$\begin{cases} s_{i+1} \leftarrow s_i0 = x_{1:i+1} \\ \kappa(s_{i+1}) \leftarrow \kappa(s_i) \end{cases}$$

We then have by definition of ν_k that $\kappa(s_{i+1}) = \nu_k(x_{1:i+1})$. The conditions of $\mathcal{I}(i+1)$ are verified.

- Case $x_{i+1} = 1$: By the invariant $I(i)$ we have $I_p \subset \Theta_k(x_{1:i})$. Moreover, by the inclusion $\Theta_k(x) \subseteq \Theta_k(x_{1:i+1})$ (by point 1 of the lemma "Construction of dyadic interval"), this gives $I_p \subset \Theta_k(x_{1:i+1})$.

  By definition of $\Theta_k(x_{1:i}1)$ and using $\mathcal{I}(i)$ and $\nu_k(x_{1:i}1) = \nu_k(x_{1:i}) + \Delta_k(x_{1:i}0)$, we have:

$$\begin{aligned} \Theta_k(x_{1:i}1) &= [\nu_k(x_{1:i}) + \Delta_k(x_{1:i}0), \nu_k(x_{1:i}) + \Delta_k(x_{1:i}0) + \Delta_k(x_{1:i}1)[\\ &= [\kappa(s_i) + \Delta_k(s_i0), \kappa(s_i) + \Delta_k(s_i0) + \Delta_k(s_i1)[\end{aligned}$$

  This is precisely condition (ii) of M_{mt}, hence its execution

$$\begin{cases} s_{i+1} \leftarrow s_i1 = x_{1:i+1} \\ \kappa(s_{i+1}) \leftarrow \kappa(s_i) + \Delta_k(s_i0) \end{cases}$$

  We then have by definition of ν_k that $\kappa(s_{i+1}) = \nu_k(x_{1:i+1})$. The invariant $\mathcal{I}(i+1)$ is verified.

Moreover, at the end of step 2 we have the writing of s_{i+1}, that is to say in this instance $x_{1:i+1}$. We therefore indeed have that the output of $M_{mt}(p)$ is a word that has x as a prefix, that is $M_{mt}(p) = x*$.

3) We must show the equality of sets $\bigcup_{p \in P_x^k} I_p = \Theta_k(x)$ by double inclusion.

$\subseteq$: Let $p \in P_x^k$. By definition of P_x^k, we directly have $I_p \subset \Theta_k(x)$. The union of the I_p is therefore contained in $\Theta_k(x)$.

$\supseteq$: The set $\Theta_k(x)$ is a semi-open interval whose endpoints are dyadic numbers (according to point 5 of the lemma "Construction of dyadic interval"). By a lemma on intervals with dyadic endpoints, there exists a finite prefix-free set $Q \subset \mathbb{B}^*$ such that:

$$\Theta_k(x) = \bigcup_{p \in Q} I_p$$

To prove the inclusion, it therefore suffices to show that for all $p \in Q$ we have $p \in P_x^k$. Let $p \in Q$. By construction of Q, we have $I_p \subset \Theta_k(x)$. This is the second condition for p to belong to P_x^k. According to point 2 of this lemma, the condition $I_p \subset \Theta_k(x)$ implies that $M_{mt}(p) = x*$. This is the first condition for p to belong to P_x^k. Both conditions being fulfilled, every $p \in Q$ is an element of P_x^k. We thus have $Q \subseteq P_x^k$. Consequently,

$$\Theta_k(x) = \bigcup_{p \in Q} I_p \subseteq \bigcup_{p \in P_x^k} I_p$$

■

Lemma: M_{mt} is monotone

Let M_{mt} be the Turing machine defined by the algorithm below for a $\phi_{lc}^{\sim}$. Then, M_{mt} is a monotone Turing machine.

Proof. 1) Output monotonicity: This is immediate. For any call $M_{mt}(p)$ at each iteration of step 2, the output s_{i+1} is formed by appending a bit to the end of s_i. The output sequence is therefore increasing with respect to the prefix relation $\leq_p$.

2) Input monotonicity: Let $p, p' \in \mathbb{B}^*$ such that $p \leq_p p'$, which implies $I_{p'} \subset I_p$. Let $(s_i)_i$ be the variables assigned in the call $M_{mt}(p)$ and let $(s'_i)_i$ be those assigned in the call $M_{mt}(p')$.

Before proving the result by loop invariance, let us show that $k = k_p$: By step 1 of M_{mt} the indices k_p and $k_{p'}$ are

$$\begin{cases} k_p = \min\{j \mid I_p \subset \Theta_j(\varepsilon)\} \\ k_{p'} = \min\{j \mid I_{p'} \subset \Theta_j(\varepsilon)\} \end{cases}$$

We thus have $I_p \subset \Theta_{k_p}$. Since $I_{p'} \subset I_p$, we also have $I_{p'} \subset \Theta_{k_p}(\varepsilon)$. The interval $I_{p'}$ is therefore contained in the intersection $\Theta_{k_p}(\varepsilon) \cap \Theta_{k_{p'}}(\varepsilon)$. As the intervals $\Theta_j(\varepsilon)$ are disjoint for distinct indices j (Previous lemma, point 2), this imposes that $k_p = k_{p'}$. Let us then set in the following $k := k_p = k_{p'}$.

Let $x = M_{mt}(p)$, thus x [is] potentially of infinite length. Let us set the invariant $\mathcal{I}(i)$ at the beginning of iteration $0 \leq i \leq \ell(x)$ of step 2:

$$\mathcal{I}(i): \quad s_i = s'_i \quad \wedge \quad I_{p'} \subset I_p \subset \Theta_k(s_i)$$

Base case ($i = 0$): By step 1 we have $s_0 = s'_0 = \varepsilon$. Furthermore, we have already shown that $k_p = k_{p'} = k$, which implies $I_p \subset \Theta_k(\varepsilon)$ and $I_{p'} \subset \Theta_k(\varepsilon)$. Given that $I_{p'} \subset I_p$, we indeed have $I_{p'} \subset I_p \subset \Theta_k(\varepsilon) = \Theta_k(s_0)$. The invariant $\mathcal{I}(0)$ is therefore verified at the beginning of step 2 at iteration $i = 0$.

Inductive step ($i \to i+1$): Assume $\mathcal{I}(i)$ is true for some $i < |x|$. This means that $M_{mt}(p)$ must necessarily write an additional bit. To do this, we must have either $I_p \subset \Theta_k(s_i 0)$ or $I_p \subset \Theta_k(s_i 1)$:

- $I_p \subset \Theta_k(s_i 0)$: Case (i) was selected by $M_{mt}(p)$, that is, the assignment $s_{i+1} \leftarrow s_i 0$ will take place. By $\mathcal{I}(i)$, we know that $I_{p'} \subset I_p$. By transitivity, $I_{p'} \subset \Theta_k(s_i 0) = \Theta_k(s'_i 0)$. Now, the intervals $\Theta_k(s_i 0)$ and $\Theta_k(s'_i 1)$ are disjoint, by a preceding lemma, thus $M_{mt}(p')$ will necessarily perform the assignment $s'_{i+1} \leftarrow s'_i 0$. We therefore indeed have $\mathcal{I}(i+1)$ true.

- $I_p \subset \Theta_k(s_i 1)$: The argument is identical. Case (ii) was selected by $M_{mt}(p)$, that is, the assignment $s_{i+1} \leftarrow s_i 1$ will take place. By $\mathcal{I}(i)$, we know that $I_{p'} \subset I_p$. By transitivity, $I_{p'} \subset \Theta_k(s_i 1) = \Theta_k(s'_i 1)$. Now, the intervals $\Theta_k(s_i 1)$ and $\Theta_k(s'_i 0)$ are disjoint, by a preceding lemma, thus $M_{mt}(p')$ will necessarily perform the assignment $s'_{i+1} \leftarrow s'_i 1$. We therefore have $\mathcal{I}(i+1)$ true.

Conclusion: By induction, for all $i < |x|$, the i-th bit of $M_{mt}(p')$ is identical to the i-th bit of $M_{mt}(p)$. This implies $M_{mt}(p) \leq_p M_{mt}(p')$. ∎

Theorem: $\Pi_{Sol} \supset \Pi_{sm}$

Assume that we have at our disposal the following elements,

- Let $\phi_{lc}^{\sim}$ be a compatible increasing dyadic lower approximator of the lower semi-computable semimeasure μ.
- a monotone Turing machine M_{mt} constructed according to the algorithm below for $\phi_{lc}^{\sim}$.

Then we have

$$\forall x \in \mathbb{B}^* \setminus \varepsilon, \quad \mu(x) = \lambda_{M_{mt}}(x)$$

Proof. For all $x \in \mathbb{B}^*$, the convergence of $\phi_{lc}^{\sim}(x, \cdot)$ to $\mu(x)$ can be written as the sum of a telescoping series:

$$\mu(x) = \lim_{i \to \infty} \phi_{lc}^{\sim}(x, i) = \sum_{k=0}^{\infty} \Delta_k(x) \quad (1)$$

By point 4 of the lemma "Construction of dyadic interval", we have $\Delta_k(x) = \lambda(\Theta_k(x))$, which yields:

$$\mu(x) = \sum_{k=0}^{\infty} \lambda(\Theta_k(x)) \quad (2)$$

The lemma "M_{mt} corresponds to Θ_k" establishes that $\Theta_k(x) = \bigcup_{p \in P_x^k} I_p$. By taking the Lebesgue measure, we obtain $\lambda(\Theta_k(x)) = \lambda\left(\bigcup_{p \in P_x^k} I_p\right)$. Substituting into (2):

$$\mu(x) = \sum_{k=0}^{\infty} \lambda\left(\bigcup_{p \in P_x^k} I_p\right) \quad (3)$$

The sets $\Theta_k(x)$ are disjoint for distinct indices k. Consequently, the union of the sets of intervals $\bigcup_{p \in P_x^k} I_p$ is also disjoint (with P_x^k defined as in the corresponding lemma). We can therefore swap the sum and the measure:

$$\mu(x) = \lambda\left(\bigcup_{k=0}^{\infty} \bigcup_{p \in P_x^k} I_p\right) = \lambda\left(\bigcup_{p \in P_x} I_p\right) \quad (4)$$

where $P_x := \bigcup_k P_x^k$ is the set of all programs for which $M_{mt}(p)$ has prefix x. By applying a lemma on "the measure of the union of dyadic intervals" to equation (4), we obtain:

$$\lambda\left(\bigcup_{p \in P_x} I_p\right) = \sum_{p' \in \lfloor P_x \rfloor} \lambda(I_{p'}) = \sum_{p' \in \lfloor P_x \rfloor} 2^{-\ell(p')} \quad (5)$$

This last sum is, by definition, the Solomonoff semimeasure $\lambda_{M_{mt}}(x)$. By relating (4) and (5), we obtain:

$$\mu(x) = \lambda_{M_{mt}}(x)$$

■

1.2.4 Universal Solomonoff Semimeasure

Definition: Universal Solomonoff Semimeasure

Let $\mathtt{U}$ be a universal monotone Turing machine. The universal Solomonoff semimeasure is the Solomonoff semimeasure associated with the machine $\mathtt{U}$,

$$\forall x \in \mathbb{B}^*, \quad \lambda_{\mathtt{U}}(x) := \sum_{\lfloor p | \mathtt{U}(p) = x* \rfloor} 2^{-\ell(p)}$$

We denote by $\mathcal{U}_{Sol}$ the set of universal Solomonoff semimeasures.

1.2.5 $\mathcal{U}_{sm} = \mathcal{U}_{Sol}$

Theorem: $\mathcal{U}_{sm}$ equals $\mathcal{U}_{Sol}$

For any universal monotone Turing machine $\mathtt{U}$ we have for all $x \in \mathbb{B}^*$ that

$$\log \mathbf{M}(x) = \log \lambda_{\mathtt{U}}(x) + O(1)$$

Proof. Let $\mathtt{U}$ be a universal monotone Turing machine.

$\leq$: We have that $\mathbf{M}$ is a lower semicomputable semimeasure. By the previous theorem there exists a monotone Turing machine M_{mt} such that for all $x \in \mathbb{B}^*$ we have $\lambda_{M_{mt}}(x) = \mathbf{M}(x)$. For I the encoding associated with $\mathtt{U}$, there exists an i such that $M_{mt}(p) = \mathtt{U}(I(i)p)$. Let us set the sets $A := \{p \mid M_{mt}(p) = x*\}$ and $B := \{I(i)p \mid \mathtt{U}(I(i)p) = x*\}$. Let us show that

$$\forall p \in \mathbb{B}^* : \quad p \in \lfloor A \rfloor \iff I(i)p \in \lfloor B \rfloor$$

$\Longrightarrow$ Consider a $p \in \lfloor A \rfloor$. By definition of a universal monotone Turing machine we have $\mathtt{U}(I(i)p) = x*$. This means that $I(i)p \in B$. It suffices to show that $I(i)p \in \lfloor B \rfloor$. For this, let us reason by contradiction and assume that $I(i)p \notin \lfloor B \rfloor$. Necessarily there exists $p' \in B$ such that $I(i)p' <_p I(i)p$. This implies $p' <_p p$. Now, knowing that $\mathtt{U}(I(i)p') = M_{mt}(p') = x*$ we have p, p' both in $\lfloor A \rfloor$, which by definition of a minimal prefix language is a contradiction. In summary $p \in \lfloor B \rfloor$.

$\Longleftarrow$ Consider a $p \in \lfloor B \rfloor$. By definition we have $\mathtt{U}(I(i)p) = M_{mt}(p) = x*$. This means $p \in A$. It suffices to show that $p \in \lfloor A \rfloor$. For this, let us reason by contradiction and assume that $p \notin \lfloor A \rfloor$. Necessarily there exists $p' \in A$ such that $p' <_p p$. This implies $I(i)p' <_p I(i)p$. Now, knowing that $\mathtt{U}(I(i)p') = M_{mt}(p') = x*$ we have $I(i)p$ and $I(i)p'$ both in $\lfloor B \rfloor$, which by definition of a minimal prefix language is a contradiction. In summary $p \in \lfloor A \rfloor$.

We can therefore write

$$\begin{aligned}
2^{-\ell(I(i))} \cdot \lambda_{M_{mt}}(x) &= 2^{-\ell(I(i))} \sum_{\underbrace{\lfloor p \mid M_{mt}(p) = x* \rfloor}_{=\lfloor A \rfloor}} 2^{-\ell(p)} \\
&= \sum_{\underbrace{\lfloor I(i)p \mid \mathtt{U}(I(i)p) = x* \rfloor}_{=\lfloor B \rfloor}} 2^{-\ell(I(i)p)} \\
&\leq \sum_{\lfloor p \mid \mathtt{U}(p) = x* \rfloor} 2^{-\ell(p)}
\end{aligned}$$

$$\leq \lambda_{\mathtt{U}}(x)$$

As $\mathbf{M}(x) = \lambda_{M_{mt}}(x)$, by setting $C = 2^{-\ell(I(i))}$ we obtain the inequality

$$C \cdot \mathbf{M}(x) \leq \lambda_{\mathtt{U}}(x)$$

$\geq$: As shown previously, every Solomonoff semimeasure is lower semicomputable, which means that $\lambda_{\mathtt{U}}$ is in Π_{sm}. Thus, as $\mathbf{M}$ is a dominant universal semimeasure, there exists $C > 0$ such that

$$\forall x \in \mathbb{B}^*, \quad \mathbf{M}(x) \geq C \cdot \lambda_{\mathtt{U}}(x)$$

■

1.3 Bayesian Mixture

1.3.1 Definition

Definition: Bayesian Mixture

Consider the following elements:

- A countable set $\mathcal{V} = \{\nu_i\}_{i=0}^{\infty}$ of lower semicomputable semimeasures for which there exists $\pi : i \in \mathcal{N} \mapsto \langle \phi_{sm,i} \rangle \in \langle \mathcal{T}_{sc} \rangle$ an effective enumeration of lower approximators satisfying,
$$\forall \nu_i \in \mathcal{V}, \quad \phi_{sm,i} \xrightarrow{lc} \nu_i$$
- A lower semicomputable function $w : \mathcal{N} \mapsto \mathbb{R}$ satisfying $w(i) > 0$ and $\sum_{i \in \mathcal{N}} w(i) \leq 1$.

We then call Bayesian Mixture the function $\xi_{w,\mathcal{V}} : \mathbb{B}^* \mapsto [0,1]$ defined for all $x \in \mathbb{B}^*$ by

$$\xi_{w,\mathcal{V}}(x) := \sum_{i \in \mathcal{N}} w(i) \cdot \nu_i(x)$$

We then denote by Π_{Bayes} the set of Bayesian Mixtures.

1.3.2 $\Pi_{Bayes} = \Pi_{sm}$

Theorem: Π_{Bayes} equals Π_{sm}

The set of Bayesian Mixtures is equal to the set of lower semicomputable semimeasures.

Proof. $\subset$: Let $\xi_{w,\mathcal{V}}$ be a Bayesian Mixture with $\phi_{sm,i} \xrightarrow{lc} \nu_i$ and $\phi_w \xrightarrow{lc} w$. Let us then set the function such that for all $\langle x, k \rangle$ in $\langle \mathbb{B}^*, \mathcal{N} \rangle$ we have

$$\Phi(x,k) = \sum_{i=0}^{k} \phi_w(i,k) \cdot \phi_{sm,i}(x,k)$$

Note that Φ is totally computable and non-decreasing. We also have that $\Phi(x,\cdot)$ is non-decreasing because $\phi_w(i,\cdot)$ and $\phi_{sm,i}(x,\cdot)$ are. Moreover $\lim_{k\to\infty} \Phi(x,k) = \xi_{w,\mathcal{V}}(x)$ for all $x \in$

$\mathbb{B}^*$. Thus we already have that $\xi_{w,\mathcal{V}}$ is lower computable. Let us now show that $\xi_{w,\mathcal{V}}(\varepsilon) \leq 1$. As ν_i is a semimeasure we have $\nu_i(\varepsilon) \leq 1$, which gives:

$$\xi_{w,\mathcal{V}}(\varepsilon) = \sum_{i \in \mathcal{N}} w(i) \cdot \nu_i(\varepsilon) \leq \sum_{i \in \mathcal{N}} w(i) \leq 1$$

We still need to show for all x in $\mathbb{B}^*$ that $\xi_{w,\mathcal{V}}(x) \geq \xi_{w,\mathcal{V}}(x0) + \xi_{w,\mathcal{V}}(x1)$. We will use the fact that ν_i is a semimeasure for this. For $x \in \mathbb{B}^*$,

$$\begin{aligned} \xi_{w,\mathcal{V}}(x) &= \sum_{i \in \mathcal{N}} w(i) \cdot \nu_i(x) \\ &\geq \sum_{i \in \mathcal{N}} w(i) \cdot [\nu_i(x0) + \nu_i(x1)] \\ &\geq \sum_{i \in \mathcal{N}} w(i) \cdot \nu_i(x0) + \sum_{i \in \mathcal{N}} w(i) \cdot \nu_i(x1) \\ &= \xi_{w,\mathcal{V}}(x1) + \xi_{w,\mathcal{V}}(x0) \end{aligned}$$

We therefore indeed have that $\xi_{w,\mathcal{V}}$ is a lower semicomputable semimeasure.

$\supset$: Let μ be a lower semicomputable semimeasure. It suffices to set for all $i \in \mathcal{N}$ the functions $w(i) = \frac{1}{2^{i+1}}$ and $\nu_i = \mu$. We then verify for any string x that,

$$\xi_{w,\mathcal{V}}(x) = \sum_{i \in \mathcal{N}} w(i) \cdot \nu_i(x) = \sum_{i \in \mathcal{N}} \frac{1}{2^{i+1}} \cdot \mu(x) = \frac{2}{1 - 1/2} \cdot \mu(x) = \mu(x)$$

■

1.3.3 Universal Bayesian Mixture

Definition: Universal Bayesian Mixture

A Bayesian Mixture $\xi_{w,\mathcal{V}}$ is said to be Universal if $\mathcal{V}$ equals the set of lower semicomputable semimeasures (i.e. $\mathcal{V} = \Pi_{sm}$). The set of Universal Bayesian Mixtures is denoted by $\mathcal{U}_{Bayes}$.

1.3.4 $\mathcal{U}_{sm} \subset \mathcal{U}_{Bayes}$

Lemma: Union of minimal languages

Let $L = \bigcup_{i \in \mathcal{N}} L_i$, where L_i are languages in $\mathbb{B}^*$. Assume that for $i, j \in \mathcal{N}$ with $i \neq j$, all $x \in L_i$ and all $y \in L_j$, any x and y are incompatible, then

$$\lfloor L \rfloor = \bigcup_{i \in \mathcal{N}} \lfloor L_i \rfloor$$

Proof. Let us proceed by double inclusion.

$\subset$: Let $x \in \bigcup_{i \in \mathcal{N}} \lfloor L_i \rfloor$. By definition of a union, there exists an index $k \in \mathbb{N}$ such that $x \in \lfloor L_k \rfloor$. To show that $x \in \lfloor L \rfloor$, we must then verify that $x \in L$ and that no proper prefix of x belongs to L.

Since $x \in \lfloor L_k \rfloor$, we have $x \in L_k$. As $L_k \subseteq L$, we deduce that $x \in L$. Suppose, for the sake of contradiction, that there exists a proper prefix $y \leq_p x$ such that $y \in L$. Since $u \in L$, there exists an index $j \in \mathcal{N}$ such that $y \in L_j$. Two cases are then possible:

- If $j = k$: We have $y \in L_k$ with $y \leq_p x$. This contradicts the hypothesis $x \in \lfloor L_k \rfloor$.
- If $j \neq k$: We have $x \in L_k$, $y \in L_j$, and $y \leq_p x$. This contradicts the hypothesis of incompatibility between the languages L_k and L_j.

The hypothesis that a proper prefix of x is in L is therefore false, that is to say that necessarily we have $x \in \lfloor L \rfloor$.

$\supset$: Let $x \in \lfloor L_i \rfloor$. Suppose by contradiction that $x \notin \lfloor L \rfloor$. Since $x \in L$, there then necessarily exists a $y \in L$ satisfying $y <_p x$. There also exists a j such that $y \in L_j$. Moreover, knowing that $x \in \lfloor L_i \rfloor$, we must have that $j \neq i$. We then have a contradiction because the assumption of the lemma states that x and y are incompatible by incompatibility of the languages L_i and L_j. In summary, we have just shown by contradiction that $x \in \lfloor L \rfloor$. ■

Theorem: $\mathcal{U}_{Sol} \subset \mathcal{U}_{Bayes}$

For every universal monotone Turing machine $\mathtt{U}$ there exists a Universal Bayesian Mixture $\xi_{w,\mathcal{V}}$ such that

$$\forall x \in \mathbb{B}^* \setminus \varepsilon, \quad \xi_{w,\mathcal{V}}(x) = \lambda_{\mathtt{U}}(x)$$

Proof. Let I be the prefix encoding associated with $\mathtt{U}$. Let $x \in \mathbb{B}^* \setminus \varepsilon$. We obtain the first equality below because if the input $e \notin \{I(i)p \mid i \in \mathcal{N}, p \in \mathbb{B}^*\}$ then $\mathtt{U}(e) = \varepsilon$, then by setting $L_i = \lfloor I(i)p \mid \mathtt{U}(I(i)\,p) = x* \rfloor$ in the preceding lemma:

$$\begin{aligned} \lfloor p \mid \mathtt{U}(p) = x* \rfloor &= \lfloor I(i)\,p \mid \mathtt{U}(I(i)\,p) = x* \rfloor \\ &= \bigcup_{i \in \mathcal{N}} \lfloor I(i)\,p \mid \mathtt{U}(I(i)\,p) = x* \rfloor \end{aligned}$$

Let us show the following equivalence for $i \in \mathcal{N}$ and the sets $A := \{I(i)\,p \mid \mathtt{U}(I(i)\,p) = x*\}$ and $B := \{p \mid M_{mt,i} = x*\}$:

$$I(i)\,p \in \lfloor A \rfloor \iff p \in \lfloor B \rfloor$$

Before showing the double implication, note that $[\dagger] : I(i)\,p \in A \iff p \in B$ and also $[\sharp] : p' <_p p \implies I(i)\,p' < I(i)\,p$.

$\implies$ Assume $I(i)\,p \in \lfloor A \rfloor$. This implies $I(i)\,p \in A$, thus $p \in B$ by $\dagger$. Suppose for the sake of contradiction that p is not minimal in B. There would exist $p' \in B$ such that $p' <_p p$.

- $p' \in B \implies I(i)\,p' \in A$ (by $\dagger$).
- $p' <_p p \implies I(i)\,p' <_p I(i)\,p$ (by $\sharp$).

This contradicts that $I(i)\,p$ is minimal in A. Thus $p \in \lfloor B \rfloor$.

$\impliedby$ Assume $p \in \lfloor B \rfloor$. This implies $p \in B$, thus $I(i)\,p \in A$ by $\dagger$. Suppose for the sake of contradiction that $I(i)\,p$ is not minimal in A. There would exist $q \in A$ such that $q <_p I(i)\,p$. Since the encoding is prefix[-free], q must be of the form $I(i)\,p'$.

- $q = I(i)\,p' \in A \implies p' \in B$ (by $\dagger$).

– $q <_p I(i)\,p \implies I(i)\,p' <_p I(i)\,p \implies p' <_p p$ (by $\sharp$).

This contradicts that p is minimal in B. Thus $I(i)\,p \in \lfloor A \rfloor$.

For all $x \in \mathbb{B}^*$ we obtain

$$\begin{aligned}
\lambda_{\mathtt{U}}(x) &= \sum_{\lfloor p | \mathtt{U}(p) = x* \rfloor} 2^{-\ell(p)} \\
&= \sum_{i \in \mathcal{N}} \sum_{\lfloor I(i)p \,|\, \mathtt{U}(I(i)\,p) = x* \rfloor} 2^{-\ell(I(i)\,p)} \\
&= \sum_{i \in \mathcal{N}} 2^{-\ell(I(i))} \sum_{\lfloor p | M_{mt,i}(p) = x* \rfloor} 2^{-\ell(p)} \\
&= \sum_{i \in \mathcal{N}} 2^{-\ell(I(i))} \cdot \lambda_{M_{mt,i}}(x)
\end{aligned}$$

The second equality comes from the union, the third from the equivalence and factoring the term $2^{-\ell(I(i))}$ out of the sum, and the last equality by $\Pi_{Sol} \supset \Pi_{sm}$ proved previously. Let us now set

$$\begin{cases} w : i \in \mathcal{N} \mapsto 2^{-\ell(I(i))} \\ \mathcal{V} = \{\lambda_{M_{mt,i}}\}_{i=0}^{\infty} \end{cases}$$

We can then verify that $\xi_{w,\mathcal{V}}$ is indeed a Universal Bayesian Mixture:

- The $M_{mt,i}$ form an enumeration of monotone Turing machines, that is $\Pi_{Sol} = \{\lambda_{M_{mt,i}}\}_{i=0}^{\infty}$. Now we have $\Pi_{Sol} = \Pi_{sm}$ therefore $\mathcal{V} = \Pi_{sm}$.
- Since $\{I(i) \mid i \in \mathcal{N}\}$ forms a prefix-free language, by Kraft's inequality $\sum_{i \in \mathcal{N}} w(i) = \sum_{i \in \mathcal{N}} 2^{-\ell(I(i))} < 1$ and moreover trivially $w(i) > 0$.

In summary, we indeed have that $\xi_{w,\mathcal{V}}$ [is] a universal Bayesian mixture such that

$$\forall x \in \mathbb{B}^* \setminus \varepsilon, \quad \xi_{w,\mathcal{V}}(x) = \lambda_{\mathtt{U}}(x)$$

■

1.3.5 $\mathcal{U}_{Sol} \supset \mathcal{U}_{Bayes}$

Lemma: Effective decomposition of semicomputable weights

Let $w : \mathcal{N} \mapsto [0,1]$ be a lower semicomputable function satisfying $\sum_{i \in \mathcal{N}} w(i) \le 1$ with $w(i) > 0$. There then exists a partially computable function $\rho : \langle \mathcal{N}, \mathcal{N} \rangle \mapsto \mathcal{N}$ such that for all $i \in \mathcal{N}$:

$$w(i) = \sum_{j=0}^{\infty} 2^{-\rho(i,j)}$$

Proof idea. Since $w(i) = \lim_{k \to \infty} M_{lc}(i,k)$, we can write $w(i)$ as a telescoping sum:

$$w(i) = M_{lc}(i,0) + (M_{lc}(i,1) - M_{lc}(i,0)) + (M_{lc}(i,2) - M_{lc}(i,1)) + \ldots$$

The algorithm computes the dyadic decomposition of each term of this sum ($\delta_k = M_{lc}(i,k) - M_{lc}(i,k-1)$) via the function Exp, then collects all the exponents. We then define $\varrho(i,j)$ as the j-th collected exponent. Thus $\sum_{j=0}^{\infty} 2^{-\rho(i,j)}$ equals $w(i)$. ■

Lemma: Effective dyadic decomposition

Let $w : \mathcal{N} \mapsto]0,1]$ be a lower semicomputable function satisfying $\sum_{i\in\mathcal{N}} w(i) \leq 1$. There exists a total computable function $\rho : \langle \mathcal{N}, \mathcal{N} \rangle \mapsto \mathcal{N}$ such that for all $i \in \mathcal{N}$:

$$w(i) = \sum_{j=0}^{\infty} 2^{-\rho(i,j)}$$

Proof. For all $\alpha = m/2^n$ a strictly positive dyadic rational number, we define the sequences $(\delta_t^\alpha)_{t\in\mathcal{N}}$ and $(s_t^\alpha)_{t\in\mathcal{N}}$ by the following recurrence: Assign $\delta_0^\alpha = \alpha$ and for all $t \in \mathcal{N}$,

1. If $\delta_t^\alpha = 0$, then $\delta_{t+1}^\alpha = 0$. In this case, s_t^α is not defined.
2. If $\delta_t^\alpha > 0$: $\begin{cases} s_t^\alpha = \min\{s \in \mathcal{N} \mid 2^{-s} \leq \delta_t^\alpha\} \\ \delta_{t+1}^\alpha = \delta_t^\alpha - 2^{-s_t^\alpha} \end{cases}$

Let us show sequentially: 1) that there exists a unique integer $N_\alpha \in \mathcal{N}$ such that $\delta_{N_\alpha+1}^\alpha = 0$ and $\delta_k^\alpha > 0$ for all $k \leq N_\alpha$; 2) the following relation is satisfied $\alpha = \sum_{k=0}^{N_\alpha} 2^{-s_k}$.

1. Note for $\delta_t^\alpha > 0$ that the relation $\delta_{t+1}^\alpha = \delta_t^\alpha - 2^{-s_t^\alpha}$ implies $\delta_{t+1}^\alpha < \delta_t^\alpha$, since $2^{-s_t^\alpha} > 0$. The sequence $(\delta_t^\alpha)_t$ is therefore strictly decreasing as long as it is positive.

   Let us prove by induction that for all $t \geq 0$, if $\delta_t^\alpha > 0$ then δ_t^α is an integer multiple of 2^{-n}. The property is true for $t = 0$. Assume it is true for a rank t, that is $\delta_t^\alpha = m_t/2^n$ for some $m_t \in \mathcal{N}, m_t \geq 1$. By the minimality of s_t^α, we have $2^{-(s_t^\alpha-1)} > \delta_t^\alpha$. Thus:

   $$2^{-(s_t^\alpha-1)} > \delta_t^\alpha = \frac{m_t}{2^n} \geq \frac{1}{2^n}$$

   The inequality $2^{-(s_t^\alpha-1)} > 2^{-n}$ implies $-(s_t^\alpha - 1) > -n$, or $s_t^\alpha - 1 < n$, and thus $s_t^\alpha \leq n$. Since $s_t^\alpha \leq n$, the integer $n - s_t^\alpha$ is non-negative. We can then write:

   $$\delta_{t+1}^\alpha = \delta_t^\alpha - 2^{-s_t^\alpha} = \frac{m_t}{2^n} - \frac{1}{2^{s_t^\alpha}} = \frac{m_t - 2^{n-s_t^\alpha}}{2^n}$$

   Let $m_{t+1} = m_t - 2^{n-s_t^\alpha}$. As m_t and $2^{n-s_t^\alpha}$ are integers, m_{t+1} is an integer. The property is thus verified at rank $t+1$, which completes the induction.

   We then have that the sequence of integers $(m_t)_{t\geq 0}$ defined by $m_t = \delta_t^\alpha \cdot 2^n$ is, as long as $m_t > 0$, strictly decreasing because $m_{t+1} = m_t - 2^{n-s_t^\alpha}$ and $2^{n-s_t^\alpha} \geq 1$. A strictly decreasing sequence of natural numbers must necessarily reach 0 in a finite number of steps. There therefore exists a first integer, denoted $N_\alpha + 1$, such that $m_{N_\alpha+1} = 0$. This implies $\delta_{N_\alpha+1}^\alpha = 0$ and that for all $k \leq N_\alpha$, $\delta_k^\alpha > 0$.

2. By summing the recurrence relation $\delta_{k+1}^\alpha - \delta_k^\alpha = -2^{-s_k^\alpha}$ for k from 0 to N_α, we obtain by telescoping:

   $$\delta_{N_\alpha+1}^\alpha - \alpha = \sum_{k=0}^{N_\alpha} (\delta_{k+1}^\alpha - \delta_k^\alpha) = -\sum_{k=0}^{N_\alpha} 2^{-s_k^\alpha}$$

   Since $\delta_{N_\alpha+1}^\alpha = 0$, we immediately deduce that $\delta_0^\alpha = \sum_{k=0}^{N_\alpha} 2^{-s_k}$.

For α in $\mathcal{Q}^{>0}$, the functions $t \in \mathcal{N} \mapsto \delta_t^\alpha$ and $t \in \{0,1,\ldots,N_\alpha\} \mapsto s_t^\alpha$ are totally computable: it suffices to apply the update rules defined above. Note that for a fixed $\delta_t^\alpha > 0$, to find s_t^α it suffices to enumerate $s = 0,1,2,\ldots$ up to the first one satisfying $2^{-s} > 0$.

Moreover $\alpha \in \mathcal{Q}^{>0} \mapsto N_\alpha$ is totally computable by the following procedure: For an input $\alpha \in \mathcal{Q}^{>0}$. Enumerate $N = 0, 1, 2, \ldots$ up to the first one satisfying $\delta^\alpha_{N+1} = 0$. Write such an N to output then accept.

By definition of a lower semicomputable function, there exists ϕ_{sc} a strict lower approximator of w, with without loss of generality $\phi_{sc}(i, 0) = 0$ for all i. Let us now set for all integers i, j the $\alpha_{i,j} := \phi_{sc}(i, j + 1) - \phi_{sc}(i, j)$. We have by closure under subtraction of dyadic values that $\alpha_{i,j}$ is dyadic. By the strict character of ϕ_{sc} we have $\alpha_{i,j} > 0$. Let us set the sequence for all i an integer S_i defined as the concatenation of the following sequences:

$$S_i := (s_t^{\alpha_{i,0}})_{t \in \{0,1,\ldots,N_{\alpha_{i,0}}\}} \frown (s_t^{\alpha_{i,1}})_{t \in \{0,1,\ldots,N_{\alpha_{i,1}}\}} \frown \cdots$$

Let us set the function $\varrho : \langle \mathcal{N}, \mathcal{N} \rangle \mapsto \mathcal{N}$ which for an input $\langle i, j \rangle$ returns the j-th element of the sequence S_i. The function ϱ is totally computable for the following effective procedure:

> For an input $\langle i, j \rangle$. By enumerating $K = 0, 1, 2, \ldots$ find the smallest K such that $\sum_{q=0}^{K} N_{\alpha_{i,q}} \leq j < \sum_{q=0}^{K+1} N_{\alpha_{i,q}}$. For $T = j - \sum_{q=0}^{K} N_{\alpha_{i,q}}$ write $s_T^{\alpha_i,K}$ to output.

We can then verify that ϱ is the function of the statement. For a given integer i:

$$\begin{aligned}
\sum_{j=0}^{\infty} 2^{-\varrho(i,j)} &= \sum_{s \in S_i} 2^{-s} \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{N_{\alpha_{i,j}}} 2^{-s_k^{\alpha_{i,j}}} \\
&= \sum_{j=0}^{\infty} \phi_{sc}(i, j+1) - \phi_{sc}(i, j) \\
&= \lim_{j \to \infty} \phi_{sc}(i, j) - 0 \\
&= w(i)
\end{aligned}$$

■

Theorem: $\mathcal{U}_{Sol} \supset \mathcal{U}_{Bayes}$

For every Universal Bayesian Mixture $\xi_{w,\mathcal{V}}$ there exists a Universal Solomonoff semimeasure $\lambda_{\mathtt{U}}$ such that

$$\forall x \in \mathbb{B}^* \setminus \varepsilon, \quad \xi_{w,\mathcal{V}}(x) = \lambda_{\mathtt{U}}(x)$$

Proof. Let $\xi_{w,\mathcal{V}}$ be a Universal Bayesian Mixture with $\mathcal{V} = \{\nu_i\}_{i=0}^{\infty}$. By definition of a Bayesian mixture, there exists an effective enumeration $\pi_{\mathcal{V}} : i \mapsto \langle \phi_{sm,i} \rangle$ such that $\phi_{sm,i} \xrightarrow{lc} \nu_i$. For each $\phi_{sm,i}$, by a theorem there exists a monotone Turing machine $M_{mt,i}$ such that $\lambda_{M_{mt,i}} = \nu_i$. We can therefore set an effective enumeration $\pi' : i \mapsto \langle M_{mt,i} \rangle$ such that $\lambda_{M_{mt,i}} = \nu_i$.

Also by definition, w is lower semicomputable and satisfies $\sum_{i \in \mathcal{N}} w(i) \leq 1$ with $w(i) > 0$. Thus by the preceding lemma there exists $\varrho : \langle i, j \rangle \mapsto k_{i,j}$ a totally computable function such that

$$w(i) = \sum_{i,j \in \mathcal{N}} 2^{-k_{i,j}}$$

By the Kraft-Chaitin theorem there exists a prefix Turing machine M and a set of strings $\{\sigma_{i,j}\}_{i,j\in\mathcal{N}}$ such that

$$\downarrow M(\sigma_{i,j}) \quad \text{with} \quad \ell(\sigma_{i,j}) = k_{i,j}$$

We then set a Turing machine $\mathtt{U}$ operating as follows

For an input $e \in \mathbb{B}^*$.
Perform the following steps in sequence:

1. Assign σ, such that $e = \sigma p$ and $\sigma \in L_\downarrow(M)$.
2. Enumerate $\langle\, i, j\,\rangle$ in $\langle \mathcal{N}, \mathcal{N}\,\rangle$ until finding $\sigma = \sigma_{i,j}$.
3. Compute $\langle\, M_{mt,i}\,\rangle \leftarrow \pi'(i)$
4. For $t = 0, 1, 2, \dots$ write to output, by appending to the end, $M_{mt,i}(p \mid t)$.

Note by construction that for all $i, j \in \mathcal{N}$ and $p \in \mathbb{B}^*$,

$$\mathtt{U}(\sigma_{i,j}p) = M_{mt,i}(p)$$

Let us now show that $\mathtt{U}$ is a universal monotone Turing machine:

- If the input e does not start with a word in $L_\downarrow(M)$ then by step 1 we have $\circlearrowleft \mathtt{U}$ with empty output, that is $\mathtt{U}(e) = \varepsilon$.
- Input monotonicity: Let $e \leq_p e'$ be two words in $\mathbb{B}^*$. If e does not start with a word in $L_\downarrow(M)$ then the assignment of step 1 fails and the output remains empty, $\mathtt{U}(e) = \varepsilon$. Otherwise, the input is of the form $e = \sigma_{i,j}p$ with $\sigma_{i,j} \in L_\downarrow(M)$. We then have by the non-ambiguity of a prefix-free language that $e' = \sigma_{i,j}p'$ with $p \leq_p p'$. Thus $\mathtt{U}(\sigma_{i,j}p) = M_{mt,i}(p)$ and $\mathtt{U}(\sigma_{i,j}p') = M_{mt,i}(p')$ which by input monotonicity of $M_{mt,i}$ gives $M_{mt,i}(p) \leq_p M_{mt,i}(p')$, that is $\mathtt{U}(e) \leq_p \mathtt{U}(e')$.
- Output monotonicity: By step 4, $\mathtt{U}(\sigma_{i,j}p)$ and $M_{mt,i}(p)$ have the same output history. As $M_{mt,i}$ is monotone this guarantees the output monotonicity of $\mathtt{U}$.

Let us now calculate the Solomonoff semimeasure $\lambda_{\mathtt{U}}$ for x in $\mathbb{B}^* \setminus \varepsilon$:

$$\begin{aligned}
\lambda_{\mathtt{U}}(x) &= \sum_{\lfloor p\,:\,\mathtt{U}(p)=x*\rfloor} 2^{-\ell(p)} \\
&= \sum_{i,j\in\mathcal{N}} \sum_{\lfloor \sigma_{i,j}p\,:\,\mathtt{U}(\sigma_{i,j}p)=x*\rfloor} 2^{-\ell(\sigma_{ij}p)} \\
&= \sum_{i,j\in\mathcal{N}} \sum_{\lfloor p\,:\,M_{mt,i}(p)=x*\rfloor} 2^{-\ell(\sigma_{ij}p)} \\
&= \sum_{i,j\in\mathcal{N}} 2^{-\ell(\sigma_{ij})} \left(\sum_{\lfloor p\,:\,M_{mt,i}(p)=x*\rfloor} 2^{-\ell(p)} \right) \\
&= \sum_{i,j\in\mathcal{N}} 2^{-k_{i,j}} \lambda_{M_{mt,i}}(x) \\
&= \sum_{i\in\mathcal{N}} \left(\sum_{j=0}^{\infty} 2^{-k_{i,j}} \right) \lambda_{M_{mt,i}}(x) \\
&= \sum_{i\in\mathcal{N}} w(i)\nu_i(x)
\end{aligned}$$

$$= \xi_{w,\mathcal{V}}(x)$$

The second equality is obtained by remarking that the $L_{i,j} = \{\sigma_{i,j}p : \mathtt{U}(\sigma_{i,j}p) = x*\}$ for $i, j \in \mathcal{N}$ form a family of incompatible languages because $\{\sigma_{i,j}\}_{i,j\in\mathcal{N}}$ is prefix-free. By applying a lemma on the union of minimal languages we obtain $\lfloor \bigcup_{i,j\in\mathcal{N}} L_{i,j} \rfloor = \bigcup_{i,j\in\mathcal{N}} \lfloor L_{i,j} \rfloor$. The third equality is obtained because for all i, j we have $\mathtt{U}(\sigma_{i,j}p) = x*$ if and only if $M_{mt,i}(p) = x*$. The rest of the equalities are straightforward. ■

1.4 Solomonoff's Induction

This part is not finished yet.

1.4.1 Posterior Probability

Definition: Posterior Probability

Let μ be a measure; we define the conditional probability of $y \in \mathbb{B}^*$ given $x \in \mathbb{B}^*$ as,

$$\mu(y \mid x) := \frac{\mu(xy)}{\mu(x)}$$

1.4.2 Entropic Inequality

Definition: Squared Hellinger Distance

Let P be a discrete probability measure ($\sum_{a\in B} P(a) = 1$) and Q a discrete semimeasure ($\sum_{a\in B} Q(a) \leq 1$), both defined on the same discrete set B. We then define the distances:

- $H(P, Q) := \sum_{a\in B} \left(\sqrt{P(a)} - \sqrt{Q(a)}\right)^2$ the Hellinger distance.
- $D(P \mid\mid Q) := \sum_{a\in B} P(a) \ln \frac{P(a)}{Q(a)}$ the Kullback-Leibler distance.

With the convention $0 \ln(0) = 0 \ln(\frac{0}{0}) = 0$.

Lemma: Entropic Inequality

Using P and Q again, we have the following inequality:

$$H(P, Q) \leq D(P \mid\mid Q)$$

Proof. For all $z \in \mathbb{R}^{>0}$ we have by concavity of the logarithm $\ln z \leq z - 1$ for all $z > 0$. For all $a \in B$ such that $P(a) > 0$, let $z = \sqrt{Q(a)/P(a)}$. The inequality becomes:

$$\ln\left(\sqrt{\frac{Q(a)}{P(a)}}\right) \leq \sqrt{\frac{Q(a)}{P(a)}} - 1$$

Let us multiply by $-2P(a)$ (which is negative), which reverses the direction of the inequality:

$$-2P(a) \ln\left(\sqrt{\frac{Q(a)}{P(a)}}\right) \geq -2P(a)\left(\sqrt{\frac{Q(a)}{P(a)}} - 1\right)$$

$$P(a)\ln\left(\left(\frac{Q(a)}{P(a)}\right)^{-1}\right) \geq -2\sqrt{P(a)Q(a)} + 2P(a)$$
$$P(a)\ln\left(\frac{P(a)}{Q(a)}\right) \geq 2P(a) - 2\sqrt{P(a)Q(a)}$$

By rearranging the terms of the Hellinger distance,

$$(\sqrt{P(a)} - \sqrt{Q(a)})^2 = P(a) - 2\sqrt{P(a)Q(a)} + Q(a) \iff$$
$$P(a) - 2\sqrt{P(a)Q(a)} = (\sqrt{P(a)} - \sqrt{Q(a)})^2 + P(a) - Q(a)$$

We thus obtain for each $a \in B$ (which is also trivially true if $P(a) = 0$):

$$P(a)\ln\left(\frac{P(a)}{Q(a)}\right) \geq \left(\sqrt{P(a)} - \sqrt{Q(a)}\right)^2 + P(a) - Q(a)$$

By summing this inequality over all $a \in B$:

$$\sum_{a\in B} P(a)\ln\frac{P(a)}{Q(a)} \geq \sum_{a\in B}\left(\sqrt{P(a)} - \sqrt{Q(a)}\right)^2 + \sum_{a\in B}(P(a) - Q(a))$$

Which is written:

$$D(P \mid\mid Q) \geq H(P,Q) + \left(\sum_{a\in B} P(a) - \sum_{a\in B} Q(a)\right)$$

Since by definition $\sum P(a) = 1$ and $\sum Q(a) \leq 1$, the last term is non-negative:

$$\sum_{a\in B} P(a) - \sum_{a\in B} Q(a) = 1 - \sum_{a\in B} Q(a) \geq 0$$

We therefore deduce the final inequality:

$$D(P \mid\mid Q) \geq H(P,Q)$$

■

1.4.3 Solomonoff Completeness

This part consists notably in using the previous entropic inequality by taking the discrete set $\mathbb{B}$. In this entire subsection we let μ be a lower semicomputable measure (not a semimeasure...) and $\mathbf{M}$ a universal semimeasure.

Lemma: Bound of the Squared Error by the KL Divergence

We define for all integers $n \geq 0$, respectively the expected squared error and the expected Kullback-Leibler divergence at step n by:

- $S_n := \sum_{x\in\mathbb{B}^{n-1}} \mu(x) \sum_{a\in\mathbb{B}} \left(\sqrt{\mathbf{M}(a|x)} - \sqrt{\mu(a|x)}\right)^2$
- $D_n := \sum_{x\in\mathbb{B}^{n-1}} \mu(x) \sum_{a\in\mathbb{B}} \mu(a|x) \ln \frac{\mu(a|x)}{\mathbf{M}(a|x)}$

We then have for all integers $n \geq 1$ that $S_n \leq D_n$.

Proof. Let $x \in \mathbb{B}^{n-1}$. Relying on the entropic inequality $H(P,Q) \leq D(P \parallel Q)$ demonstrated previously:

$$\sum_{a\in\mathbb{B}} \left(\sqrt{P(a)} - \sqrt{Q(a)}\right)^2 \leq \sum_{a\in\mathbb{B}} P(a) \ln \frac{P(a)}{Q(a)}$$

By setting $P(a) = \mu(a|x)$ and $Q(a) = \mathbf{M}(a|x)$ for a fixed x, we obtain:

$$\sum_{a\in\mathbb{B}} \left(\sqrt{\mu(a|x)} - \sqrt{\mathbf{M}(a|x)}\right)^2 \leq \sum_{a\in\mathbb{B}} \mu(a|x) \ln \frac{\mu(a|x)}{\mathbf{M}(a|x)}$$

Let us multiply this inequality by the probability $\mu(x) \geq 0$ and sum over all histories $x \in \mathbb{B}^{n-1}$:

$$\sum_{x\in\mathbb{B}^{n-1}} \mu(x) \sum_{a\in\mathbb{B}} \left(\sqrt{\mu(a|x)} - \sqrt{\mathbf{M}(a|x)}\right)^2 \leq \sum_{x\in\mathbb{B}^{n-1}} \mu(x) \sum_{a\in\mathbb{B}} \mu(a|x) \ln \frac{\mu(a|x)}{\mathbf{M}(a|x)}$$

By definition, the left term is S_n (the expected squared Hellinger distance) and the right one is D_n. The inequality $S_n \leq D_n$ is thus established. ■

Lemma: Bound of the Sum of KL Divergences

The series of expected KL divergences is finite and bounded by:

$$\sum_{n=1}^{\infty} D_n < \infty$$

Proof. Starting from the definition of D_n, for all $m \geq n$:

$$\begin{aligned}
&\sum_{x_{1:n}\in\mathbb{B}^n} \mu(x_{1:n}) \ln\left(\frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})}\right) \\
&= \sum_{x_{1:n}\in\mathbb{B}^n} \left(\sum_{x_{n+1:m}\in\mathbb{B}^{m-n}} \mu(x_{1:n}x_{n+1:m})\right) \ln\left(\frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})}\right) \\
&= \sum_{x_{1:n}\in\mathbb{B}^n} \sum_{x_{n+1:m}\in\mathbb{B}^{m-n}} \mu(x_{1:n}x_{n+1:m}) \ln\left(\frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})}\right) \\
&= \sum_{x_{1:m}\in\mathbb{B}^m} \mu(x_{1:m}) \ln\left(\frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})}\right)
\end{aligned}$$

The first equality follows from the law of total probability $\mu(x_{1:n}) = \sum_{y\in\mathbb{B}^{m-n}} \mu(x_{1:n}y)$ where $x_{1:n} \in \mathbb{B}^n$ is fixed. The rest of the derivation consists of reorganizing the sums.

Starting from the sum of the D_n, we have:

$$\begin{aligned}
\sum_{n=1}^{m} D_n &= \sum_{n=1}^{m} \left[\sum_{x_{1:m}\in\mathbb{B}^m} \mu(x_{1:m}) \ln\left(\frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})}\right)\right] \\
&= \sum_{x_{1:m}\in\mathbb{B}^m} \sum_{n=1}^{m} \mu(x_{1:m}) \ln\left(\frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})}\right) \\
&= \sum_{x_{1:m}\in\mathbb{B}^m} \mu(x_{1:m}) \left[\sum_{n=1}^{m} \ln\left(\frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})}\right)\right]
\end{aligned}$$

$$= \sum_{x_{1:m} \in \mathbb{B}^m} \mu(x_{1:m}) \ln \left(\prod_{n=1}^{m} \frac{\mu(x_n|x_{1:n-1})}{\mathbf{M}(x_n|x_{1:n-1})} \right)$$

$$= \sum_{x_{1:m} \in \mathbb{B}^m} \mu(x_{1:m}) \ln \left(\frac{\mu(x_{1:m})}{\mathbf{M}(x_{1:m})} \right)$$

The last equality is justified by the chain rule (or telescoping product) for (semi-)measures. For any sequence $x_{1:m}$, by defining $x_{1:0}$ as the empty sequence ε:

$$\prod_{n=1}^{m} \mu(x_n|x_{1:n-1}) = \mu(x_1|\varepsilon) \cdot \mu(x_2|x_1) \cdots \mu(x_m|x_{1:m-1}) = \mu(x_{1:m})$$

and identically for $M(x_{1:m})$. By definition of a universal semimeasure there exists $c > 0$ an integer constant such that $\forall x \in \mathbb{B}^*, \mathbf{M}(x) \geq c \cdot \mu(x)$. We thus obtain

$$\sum_{n=1}^{m} D_n \leq \sum_{x_{1:m} \in \mathbb{B}^m} \mu(x_{1:m}) \ln \left(\frac{\mu(x_{1:m})}{\mathbf{M}(x_{1:m})} \right) \leq \ln \left(\frac{1}{c} \right) \sum_{x_{1:m} \in \mathbb{B}^m} \mu(x_{1:m}) < \infty$$

■

Theorem: Solomonoff Completeness

Let μ be a lower semicomputable measure. The total sum of expected squared errors is finite and bounded:

$$\sum_{t=1}^{\infty} \sum_{x \in \mathbb{B}^{t-1}} \mu(x) \left(\sqrt{\mathbf{M}(0|x)} - \sqrt{\mu(0|x)} \right)^2 < \infty$$

Proof. By the lemma above $S_n \leq D_n$ for all $n \geq 1$. By summing this inequality from $n = 1$ to infinity, we obtain:

$$\sum_{n=1}^{\infty} S_n \leq \sum_{n=1}^{\infty} D_n$$

The previous Lemma establishes that the series on the right converges; consequently $\sum_{n=1}^{\infty} S_n < \infty$. Consequently by making S_n explicit:

$$\sum_{t=1}^{\infty} \sum_{x \in \mathbb{B}^{t-1}} \mu(x) \left(M(0|x) - \mu(0|x) \right)^2 < \infty$$

■